

(21) Application No 9024835.2

(22) Date of filing 15.11.1990

(71) Applicant  
Sony Broadcast & Communications Limited

(Incorporated in the United Kingdom)

Jays Close, Viables, Basingstoke, Hampshire,  
RG22 4SB, United Kingdom

(72) Inventors  
Clive Henry Gillard  
James Edward Burns

(74) Agent and/or Address for Service  
D Young & Co  
10 Staple Inn, London, WC1V 7RD, United Kingdom

(51) INT.CL<sup>5</sup>  
H04N 7/01 // H04N 5/253 5/87

(52) UK CL (Edition K)  
H4F FD1B9 FD1D1 FD10 FD12K9 FD12M FD12X  
FD30G FD30K FD30T1 FD30T3 FD30X FD54 FGXX

(56) Documents cited  
None

(58) Field of search  
UK CL (Edition K) H4F FEP FER FGXX  
INT CL<sup>5</sup> G06F, H04N  
Online databases: WPI

(54) Motion compensated interpolation of images

(57) For each pixel location  $(x, y)$  in an output image which is to be interpolated half-way between a pair of input images, a respective motion vector  $(m, n)$  is produced indicative of the estimated motion of that pixel between the two images, and the value of the output pixel is derived from the determined values of the pixels at locations  $(x - m/2 - 1/4, y - n/2 - 1/4)$  and  $(x + m/2 - 1/4, y + n/2 - 1/4)$  in the respective input images. Those values are determined by spatial interpolation in the input images, and the  $(-1/4, -1/4)$  offset ensures equal magnitude responses of the spatial interpolation operation irrespective of whether  $m$  and  $n$  are odd or even numbers. In the case where the output image is temporally aligned with an input image, each pixel in the output image is interpolated between the pixels at locations  $(x - 1/4, y - 1/4)$  in the temporally aligned input image and  $(x + m - 1/4, y + n - 1/4)$  in the succeeding input image in order to reduce noise modulation in the output image.

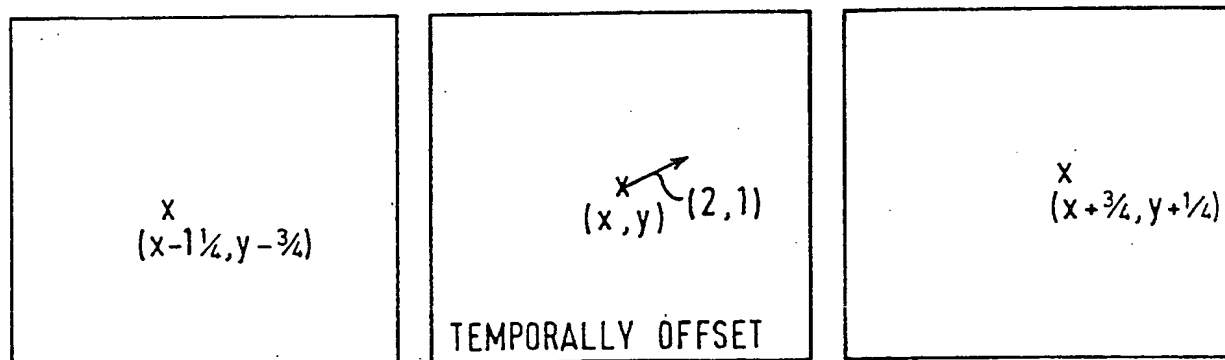


FIG. 63C.

1/45

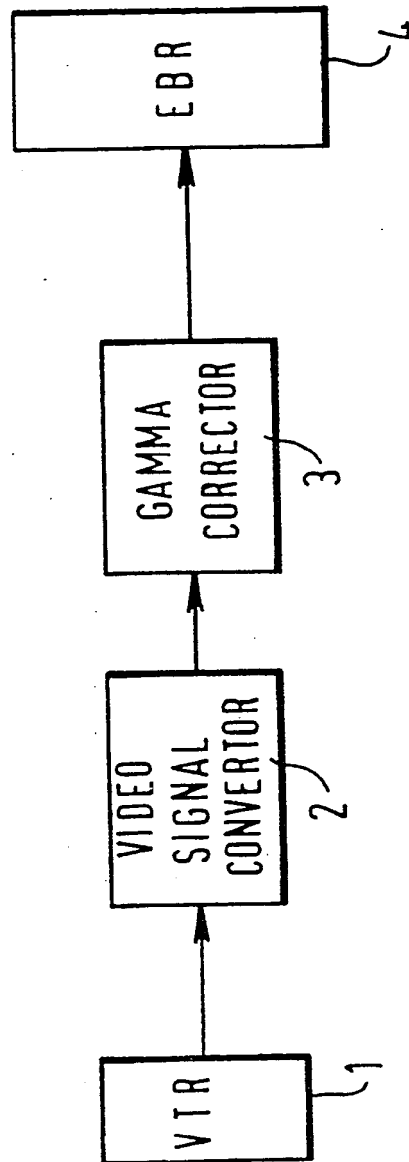


FIG. 1.

2/45

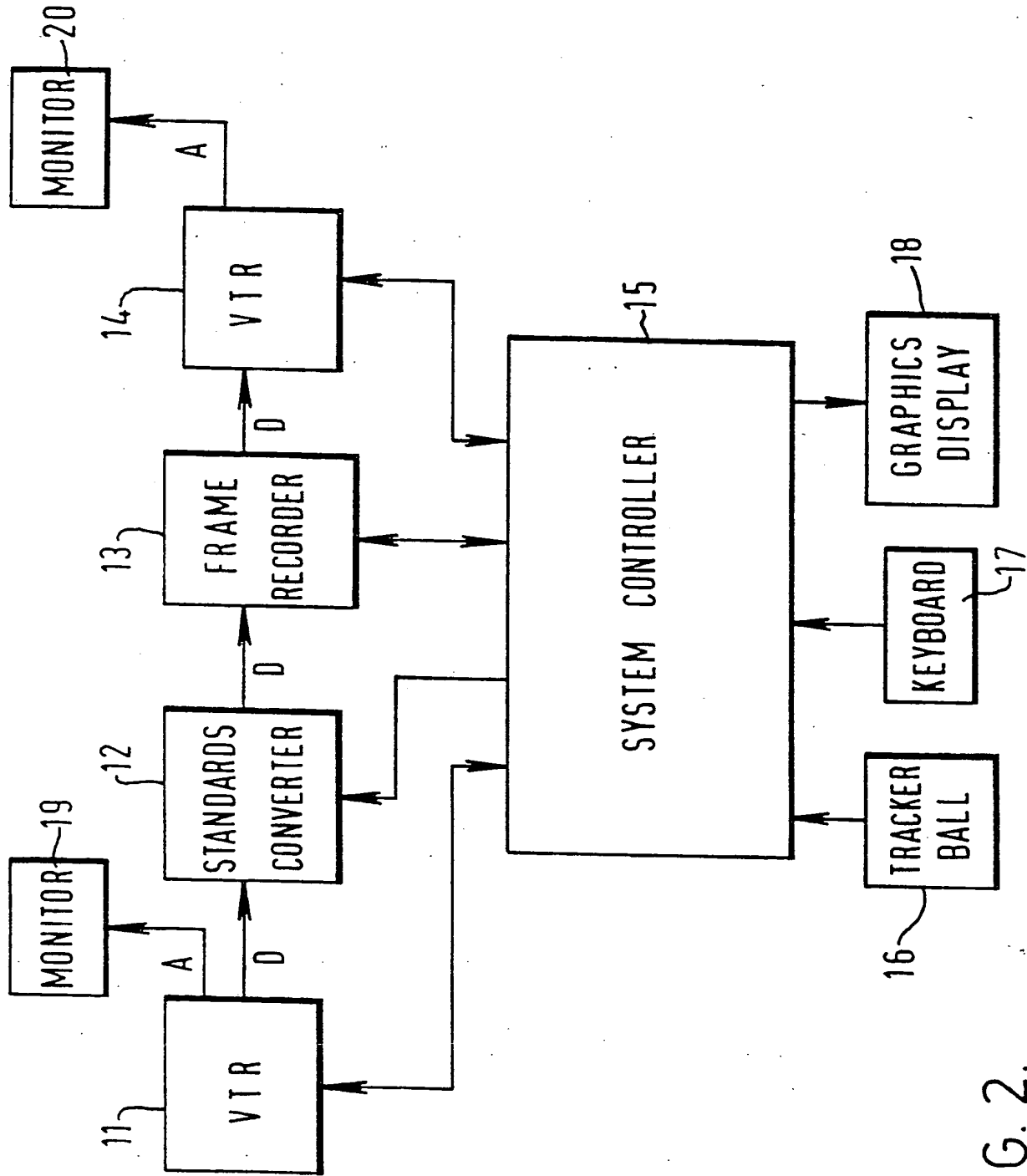


FIG. 2.

3/45

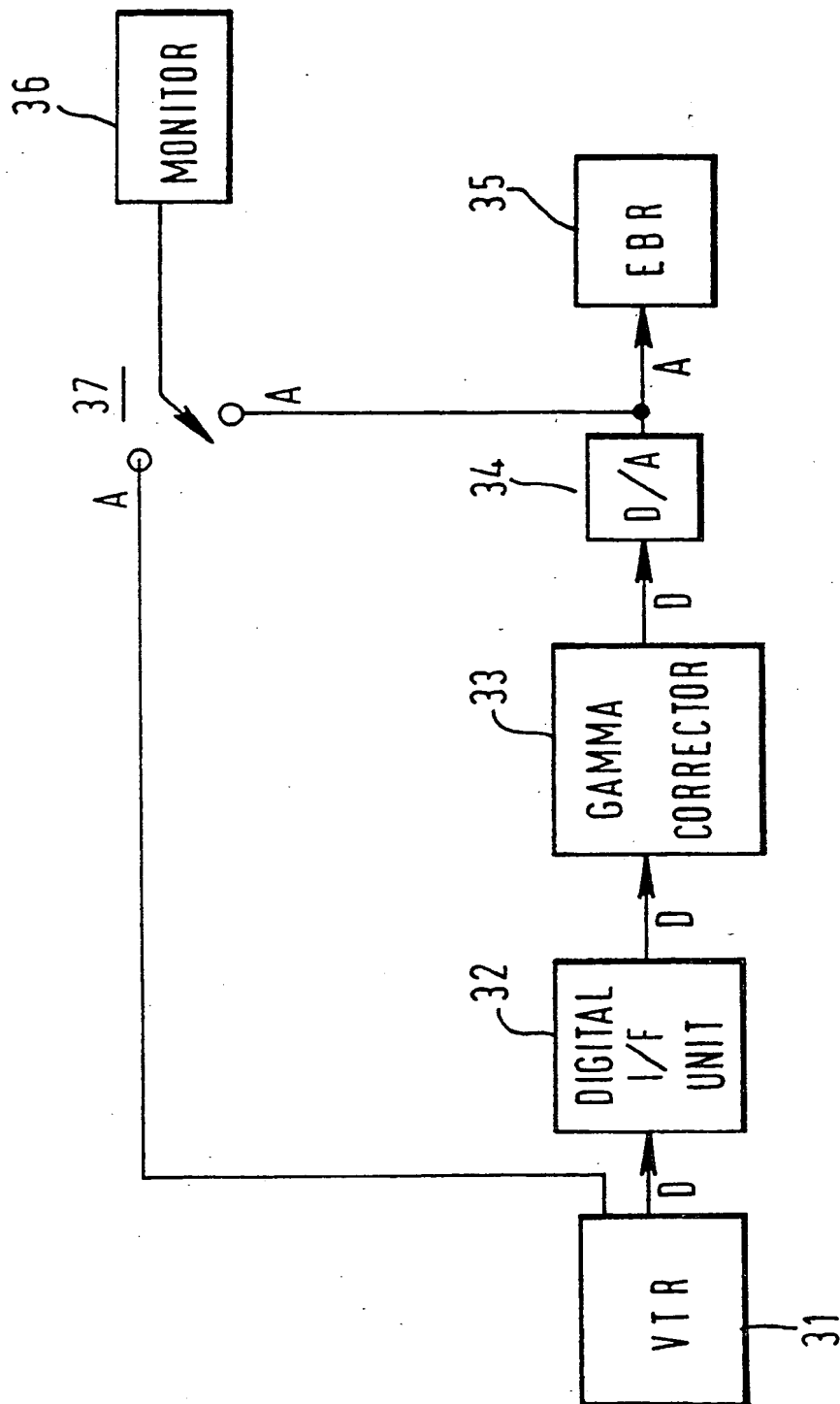


FIG. 3.

4145

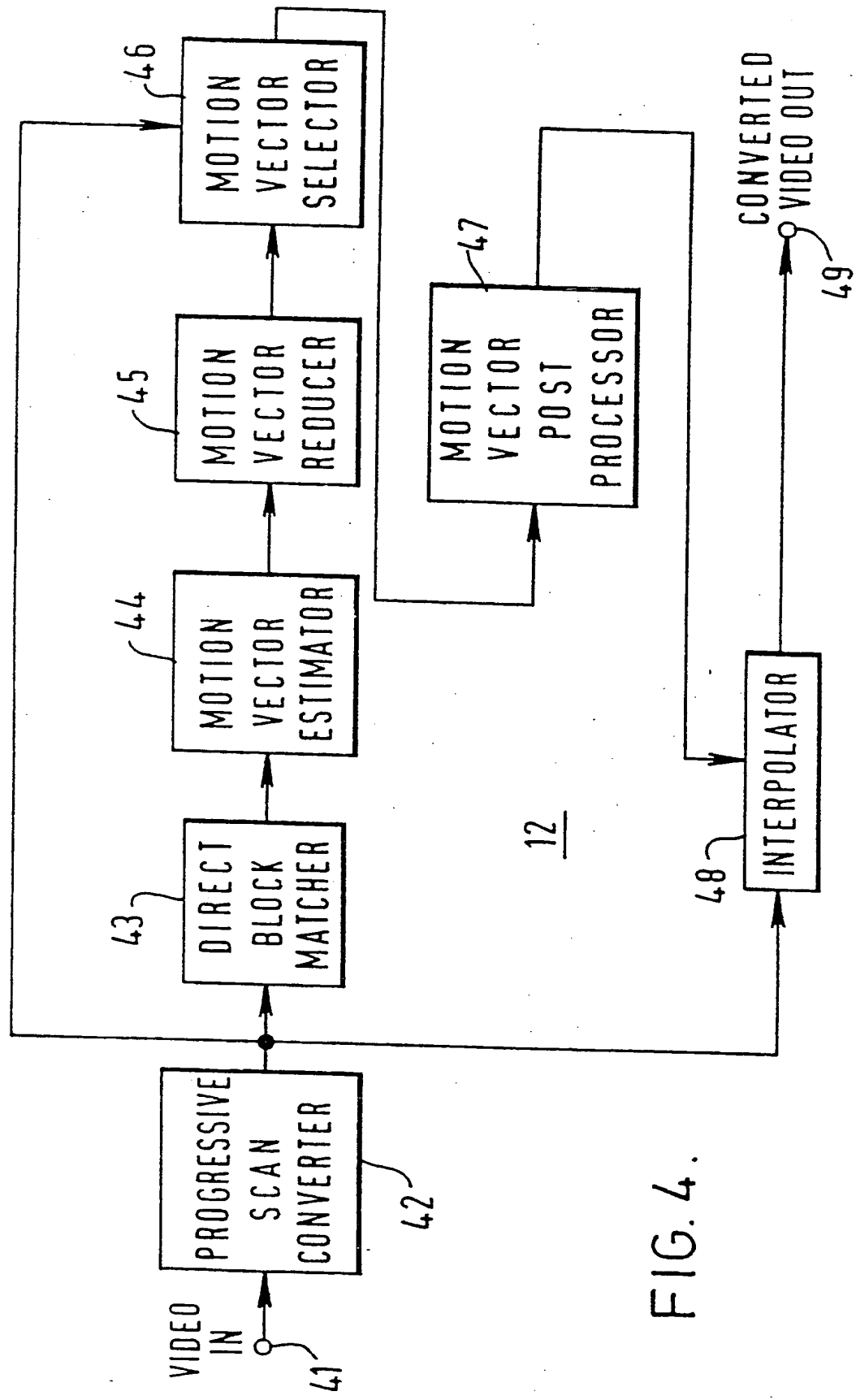
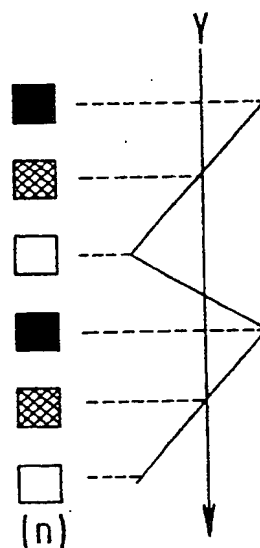
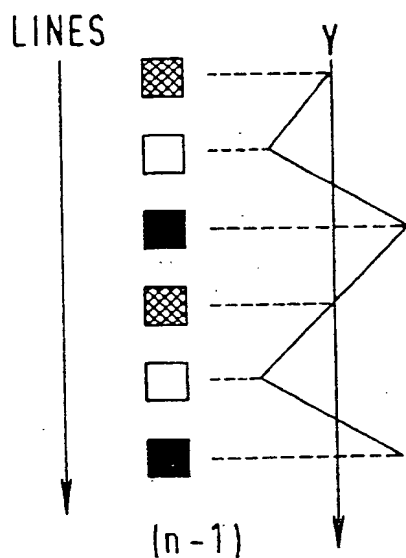
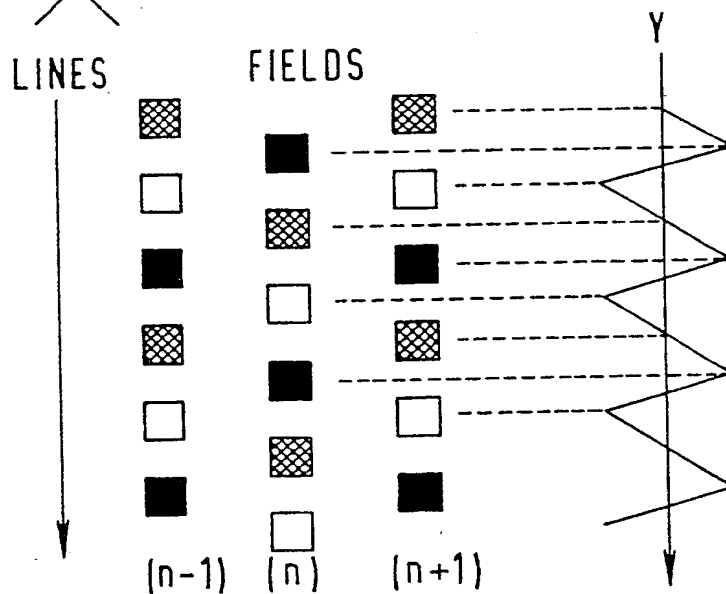
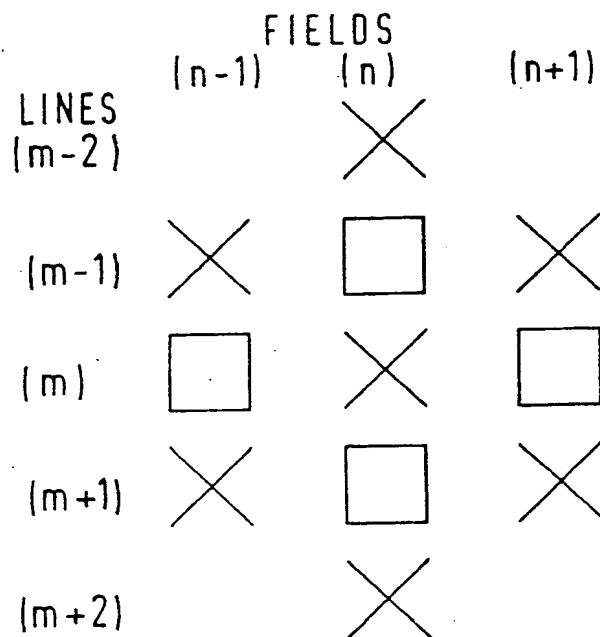


FIG. 4.

○



6/45

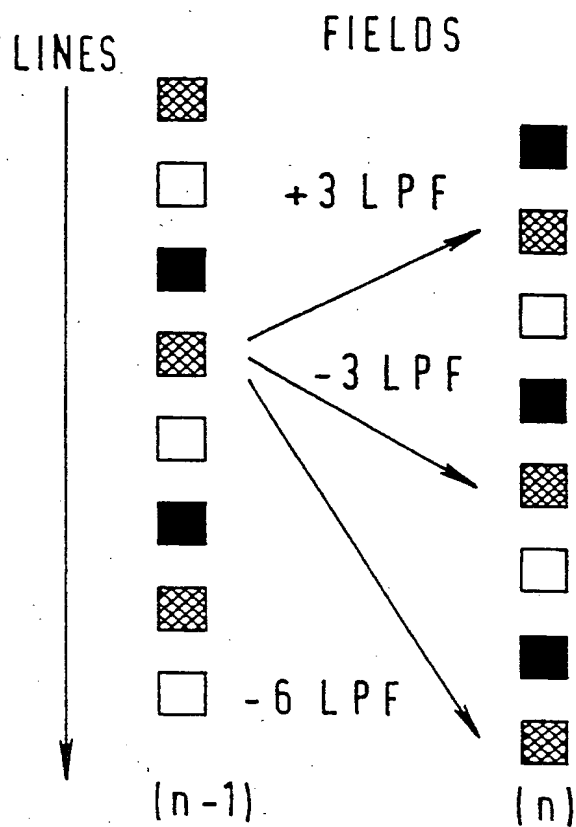


FIG. 8.

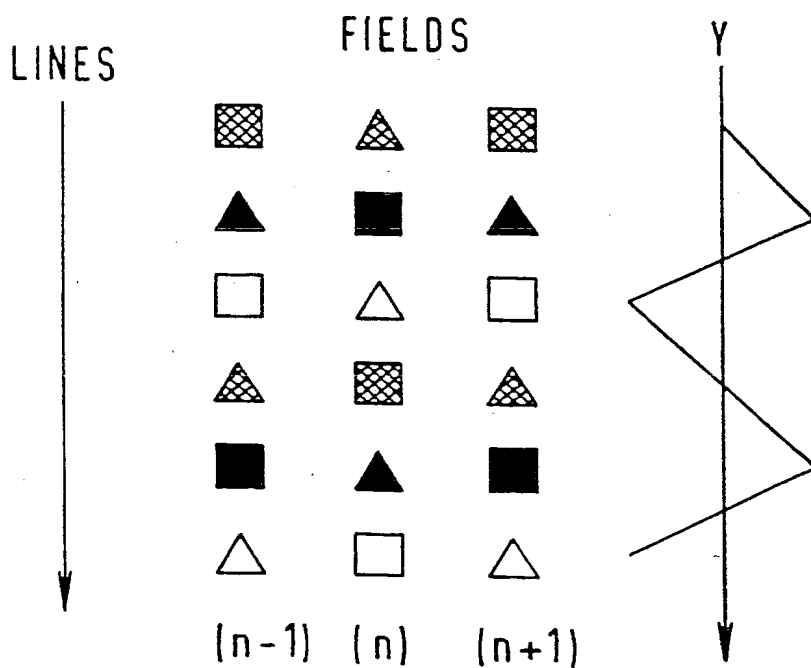
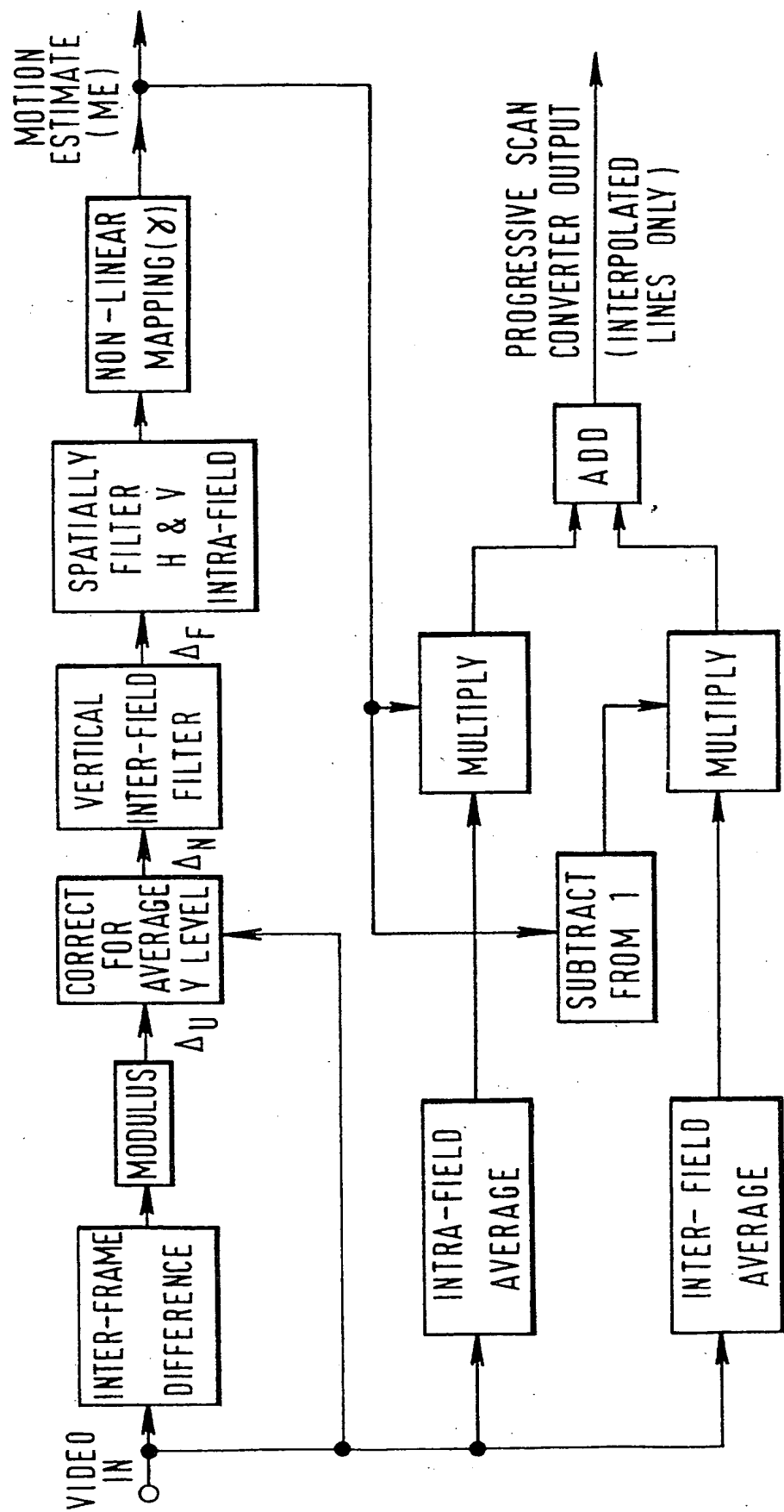


FIG. 9.

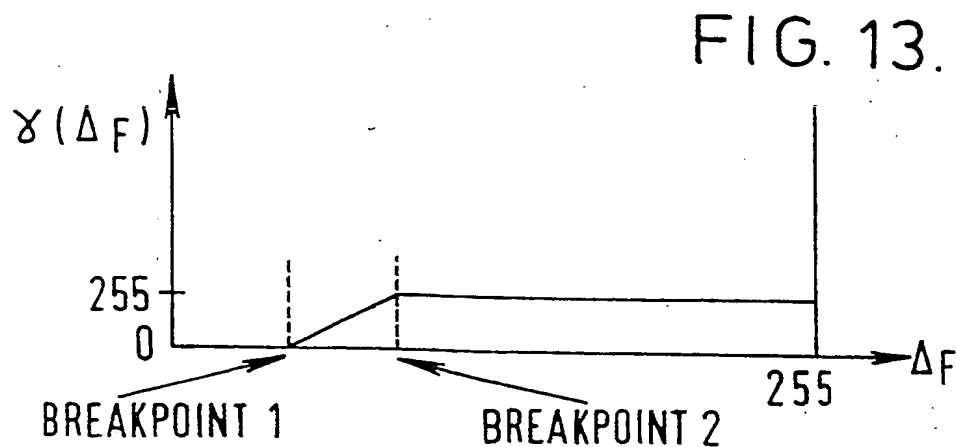
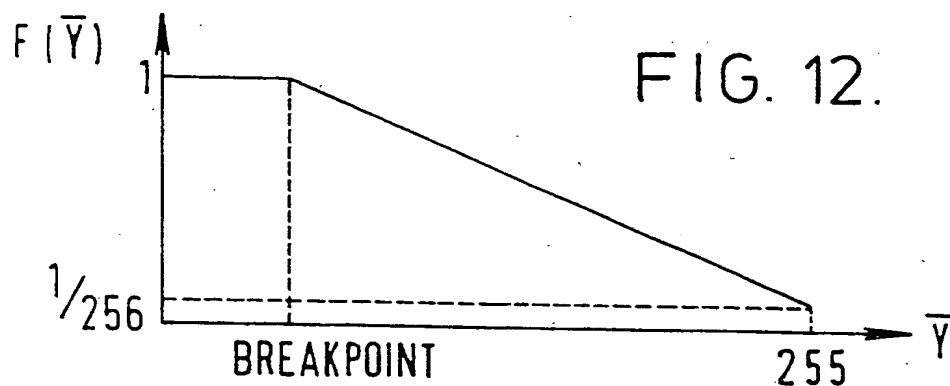
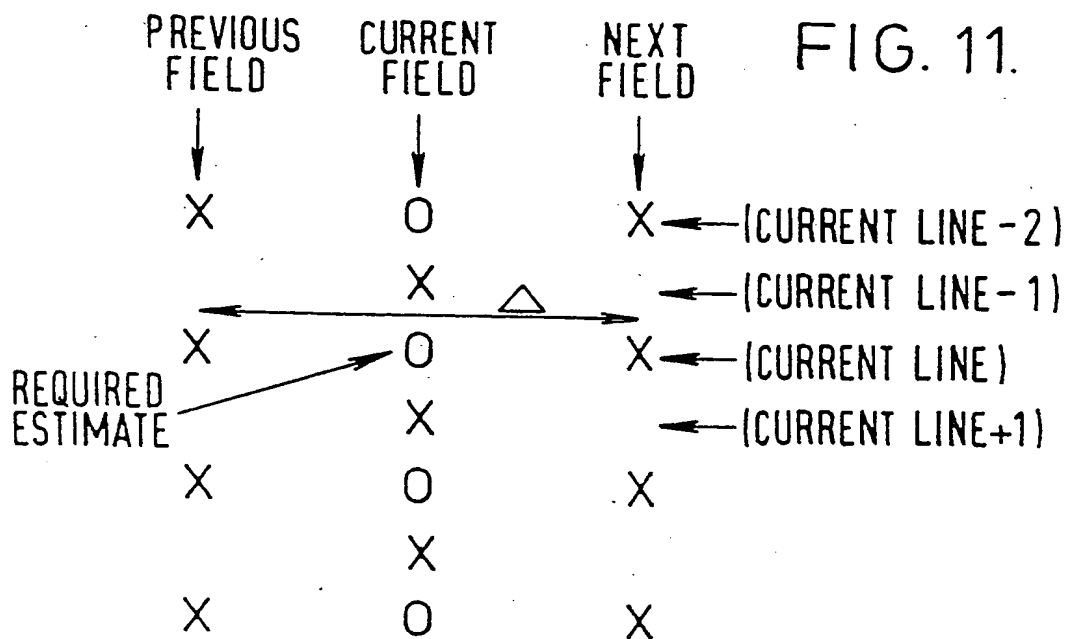
7/45

FIG.10.





8/45



9/45

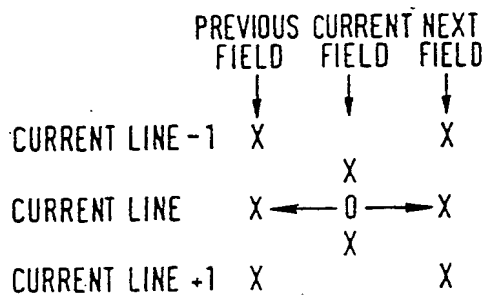
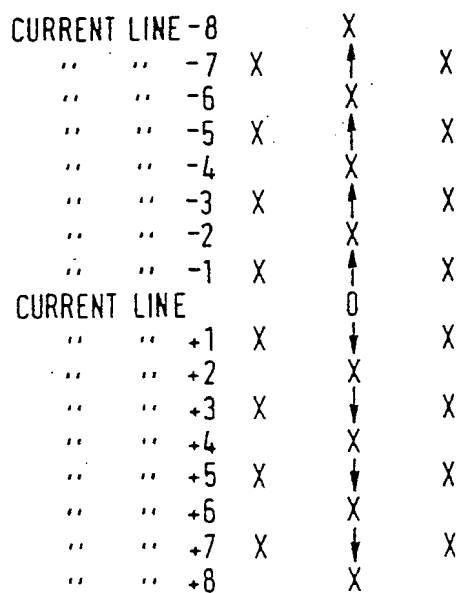


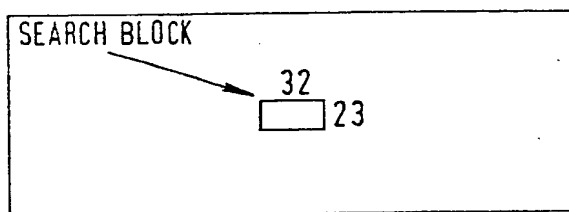
FIG. 14.

INTER-FRAME AVERAGE  
FOR STATIC PICTURES

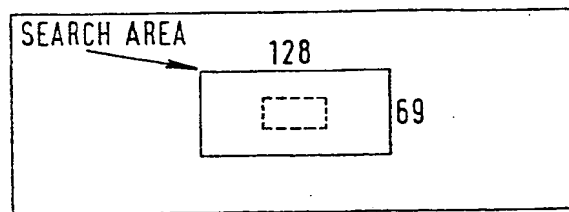


INTRA-FIELD INTERPOLATION  
FOR MOVING PICTURES

FIG. 15.



FRAME N



FRAME (N+1)

10/45

FIG. 16.

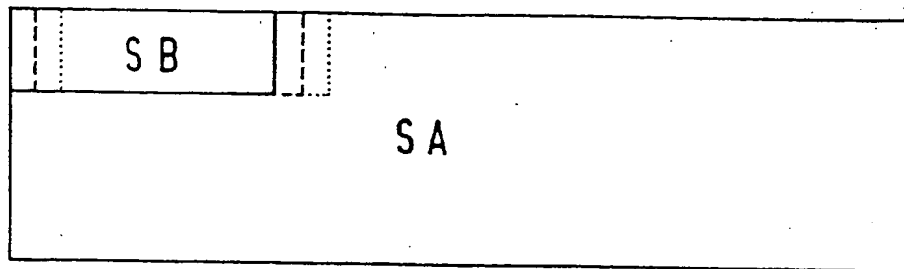


FIG. 17.

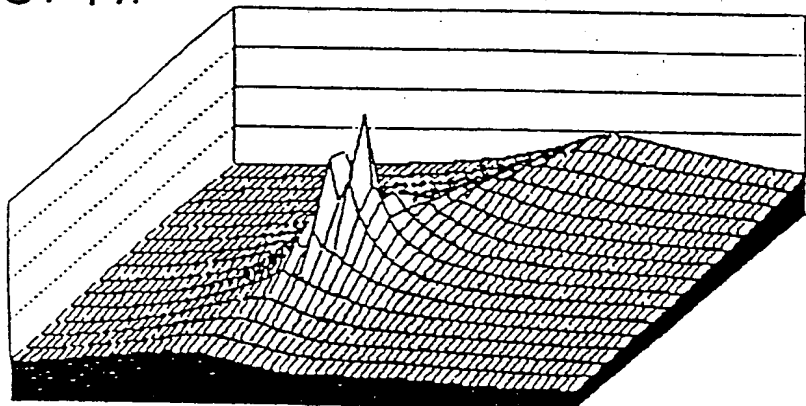
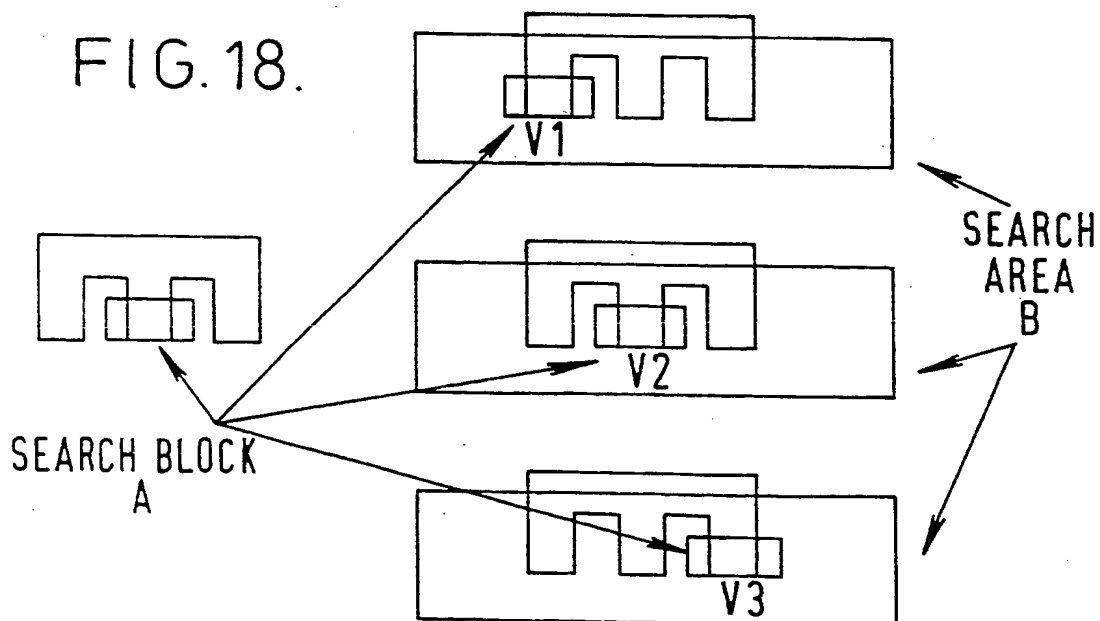
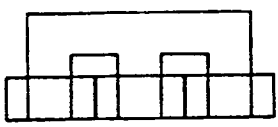


FIG. 18.

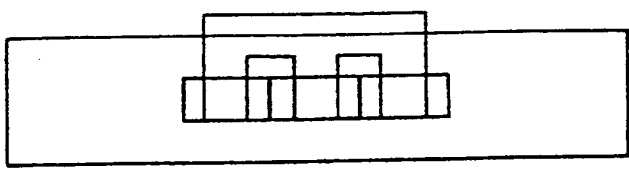


11/45

FIG. 19.



SEARCH BLOCK 3A



SEARCH AREA

FIG. 20.

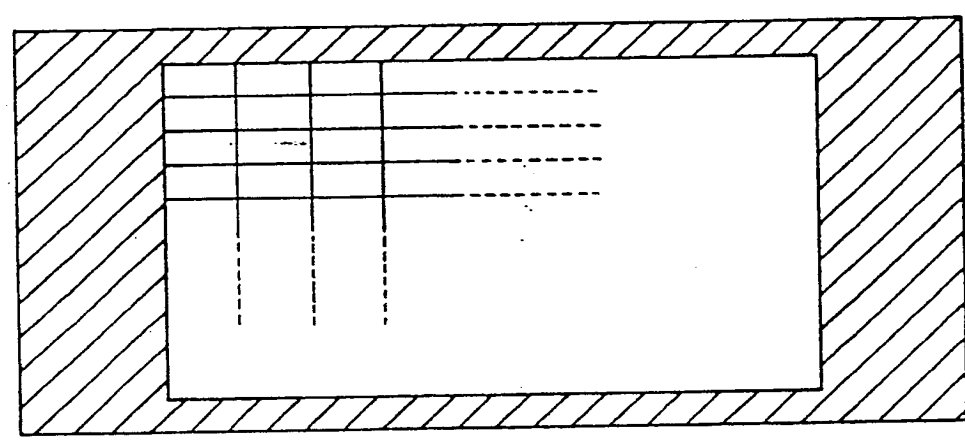
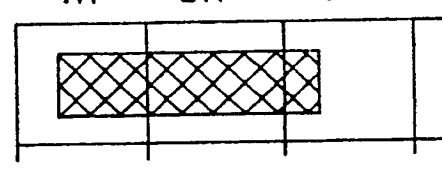


FIG. 21.

SEARCH BLOCKS.

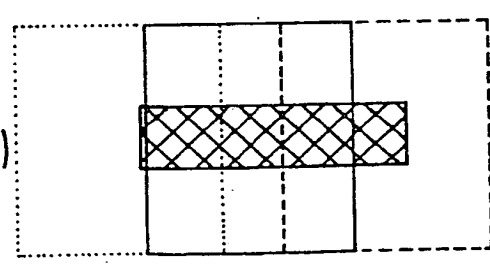
1A 2A 3A

FRAME (t)



MOTION (5, 0)

FRAME (t+1)



1B 2B 3B

SEARCH AREAS.

12/45

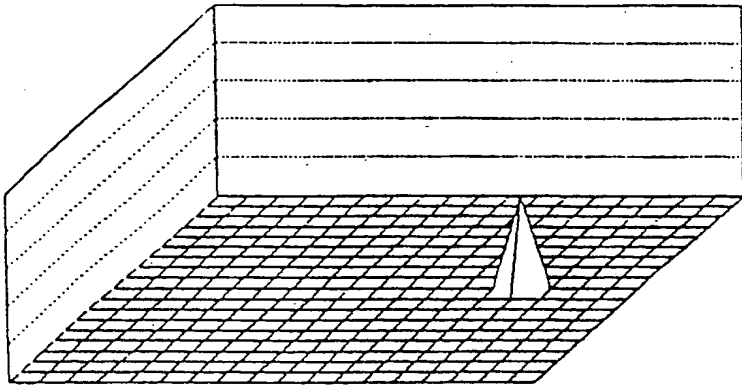


FIG. 22.

FIG. 23.

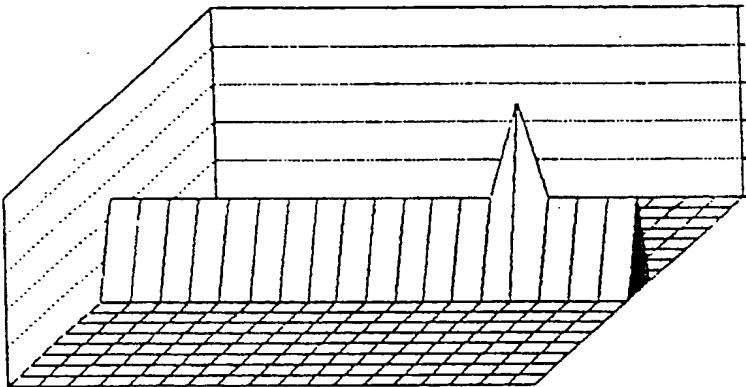
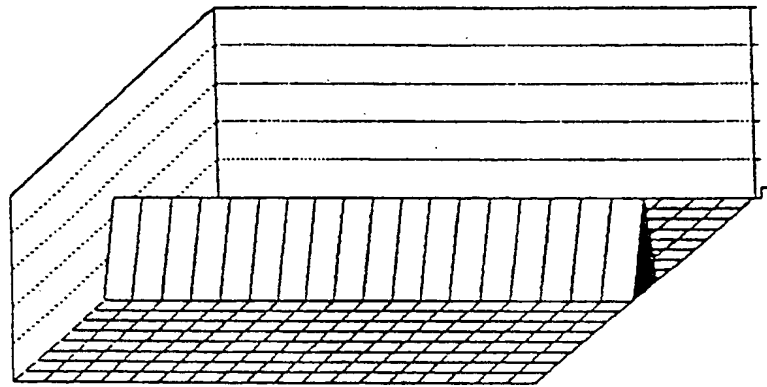
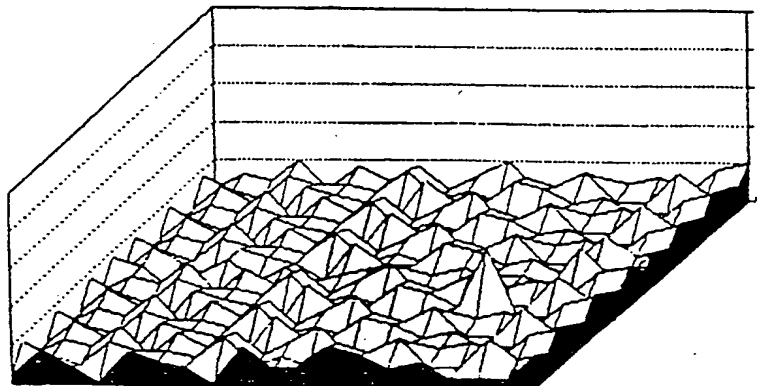


FIG. 24.

FIG. 25.



13/45

FIG. 26.

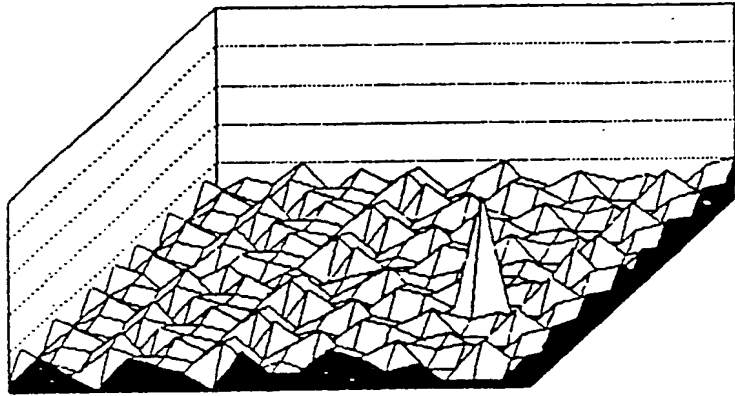


FIG. 27.

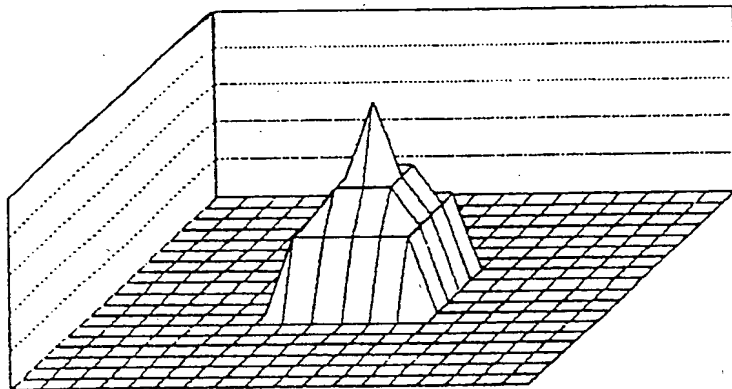
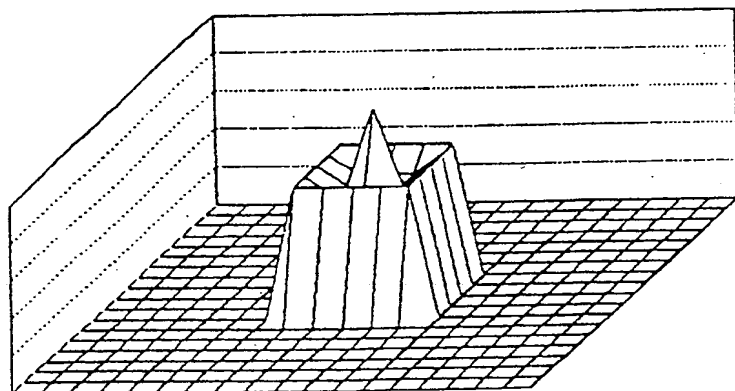
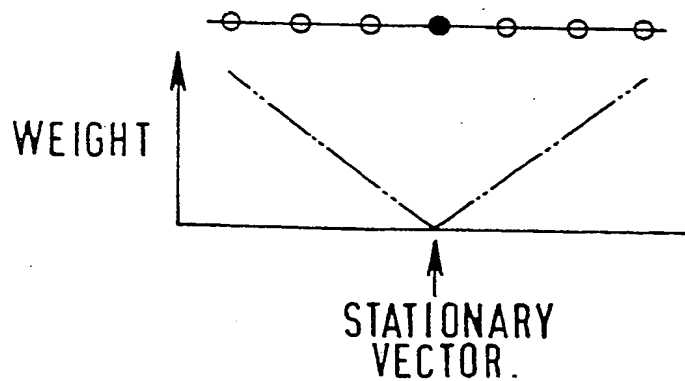
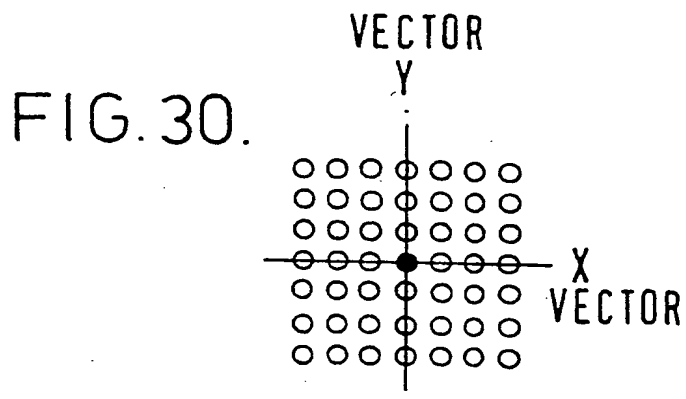
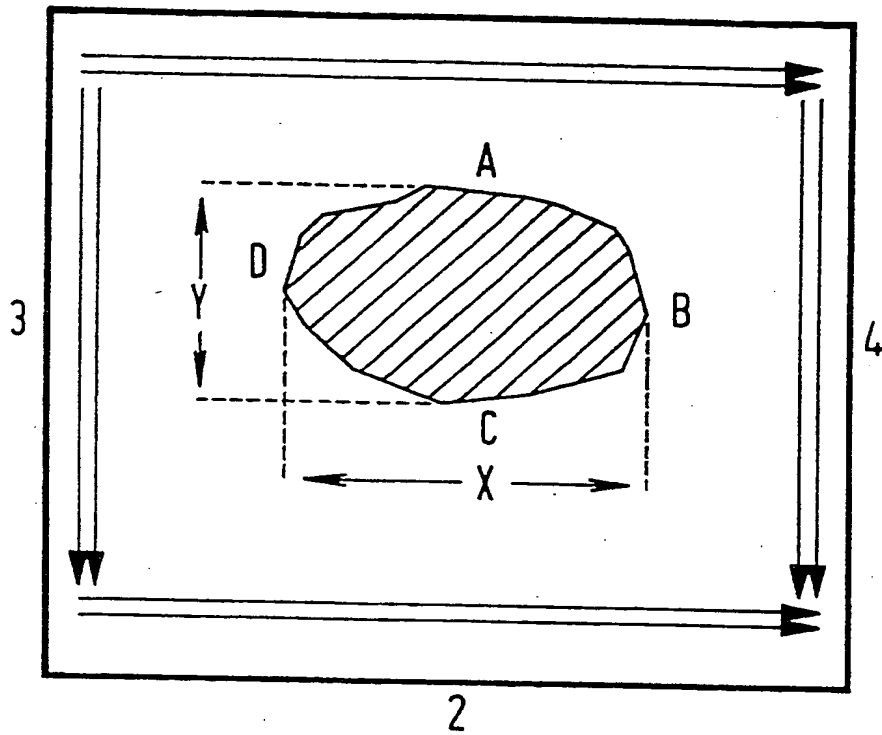


FIG. 28.



1/4/45

FIG. 29. 1



15/45

FIG. 31. FRAME OF VIDEO

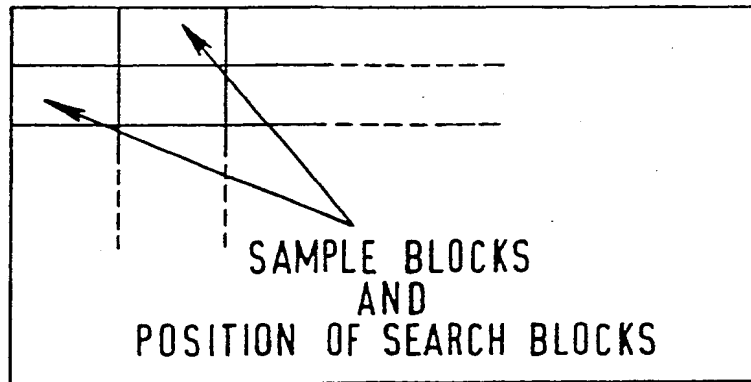
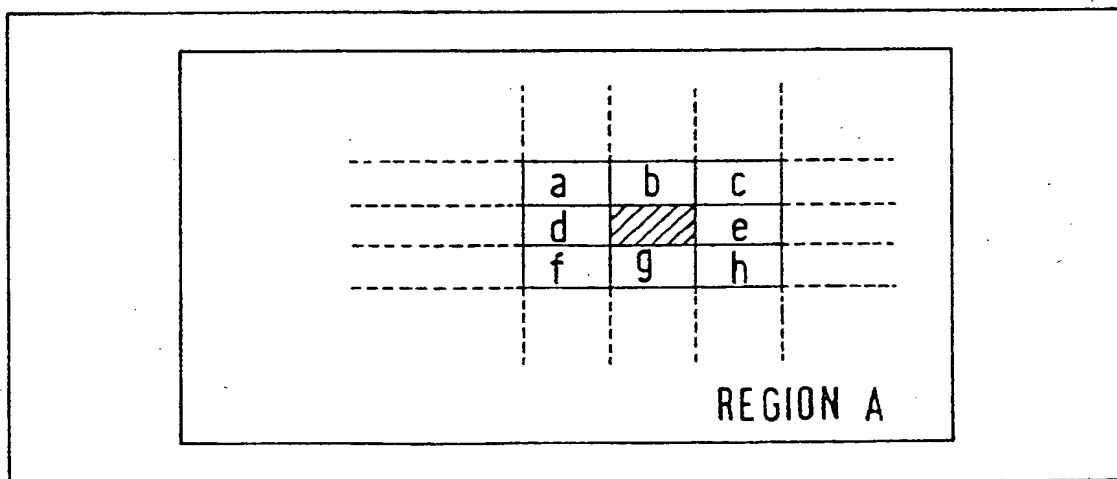


FIG. 32.

C	C	C	C	C	C						
C	C	B	B	B	B						
C	C	B	A	A	A						
C	C	B	A	A	A						
C	C	B	A	A	A						
C	C	B	A	A	A						

FIG. 33.





16/45

FIG. 34.

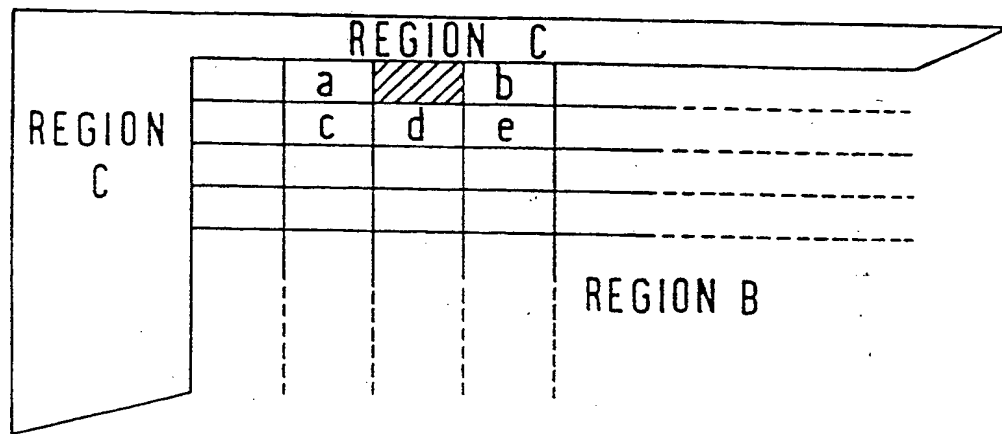
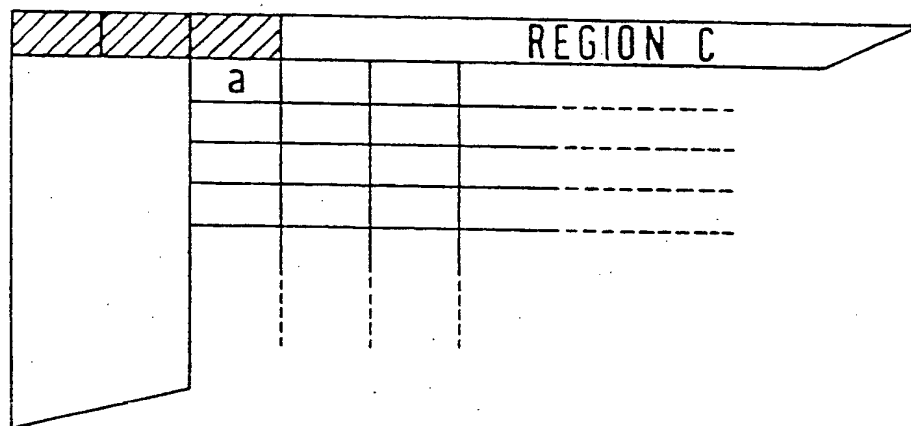


FIG. 35.



17/45

FIG.36.

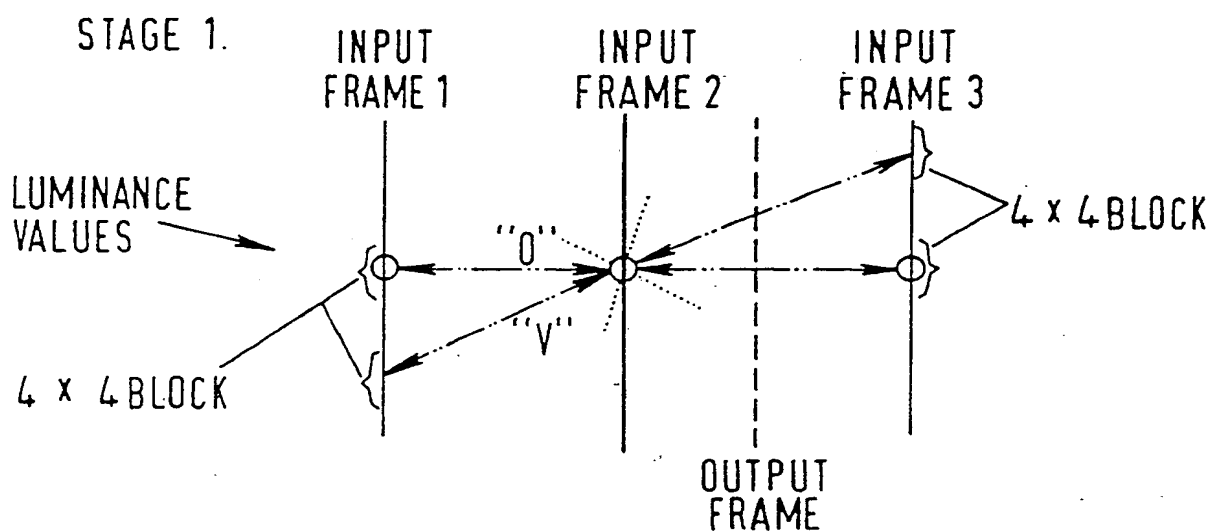
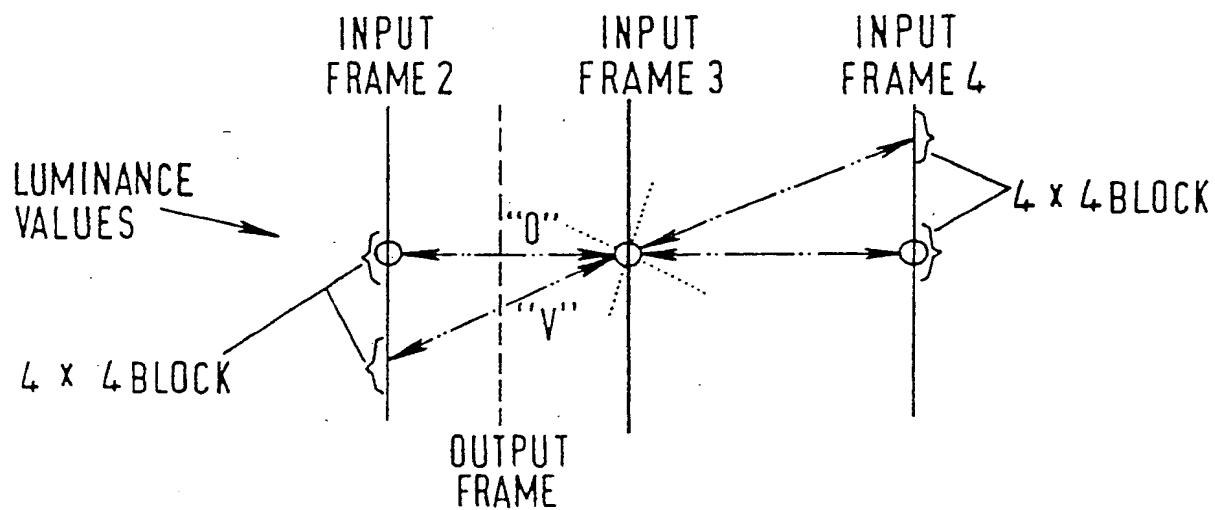


FIG.37.





19/45

FIG. 41.

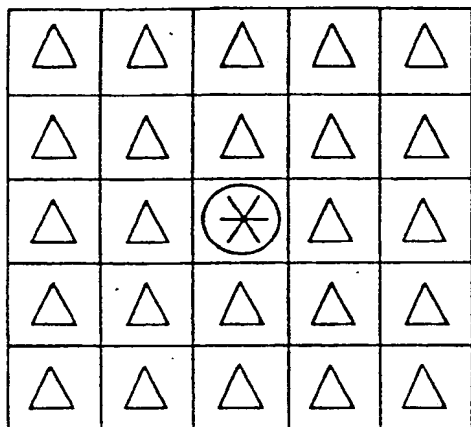


FIG. 42.

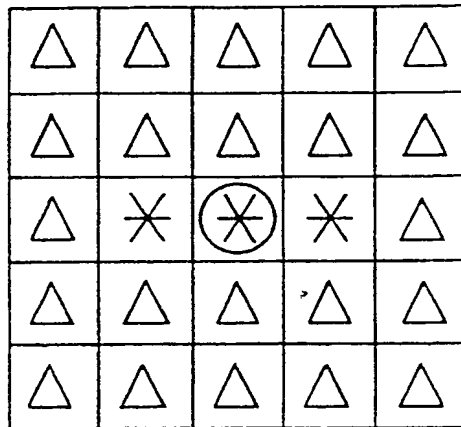


FIG. 43.

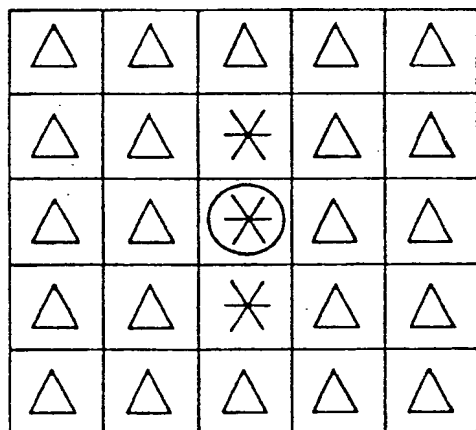
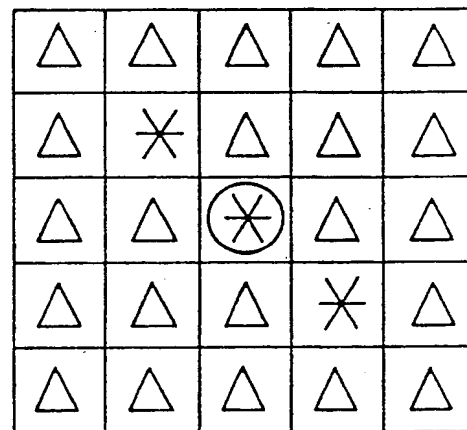


FIG. 44.



20/45

FIG. 45.

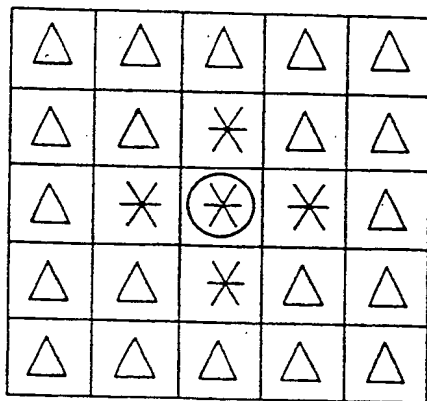


FIG. 46.

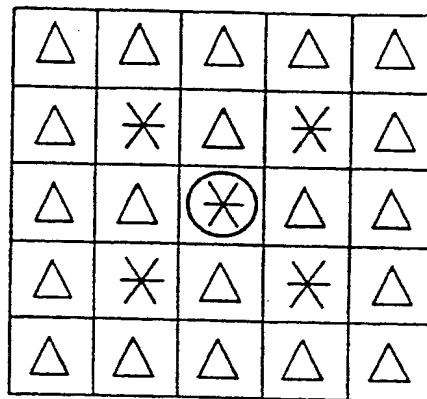
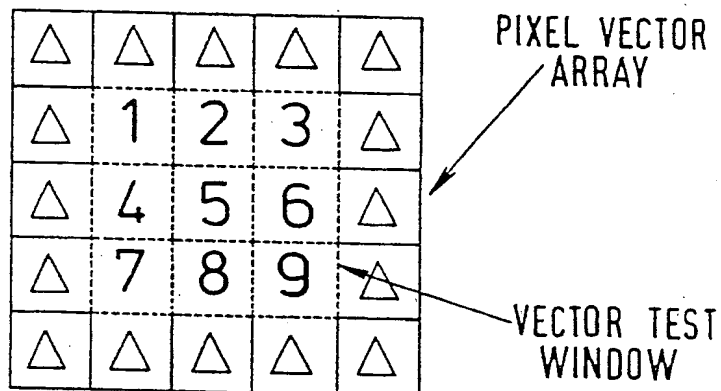


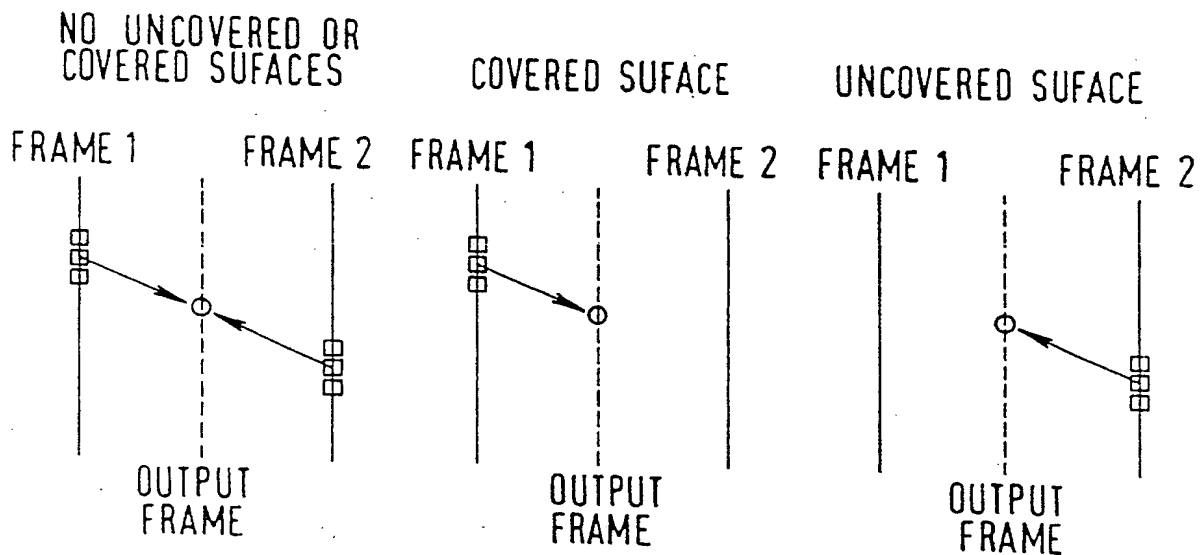
FIG. 47.



LEGEND

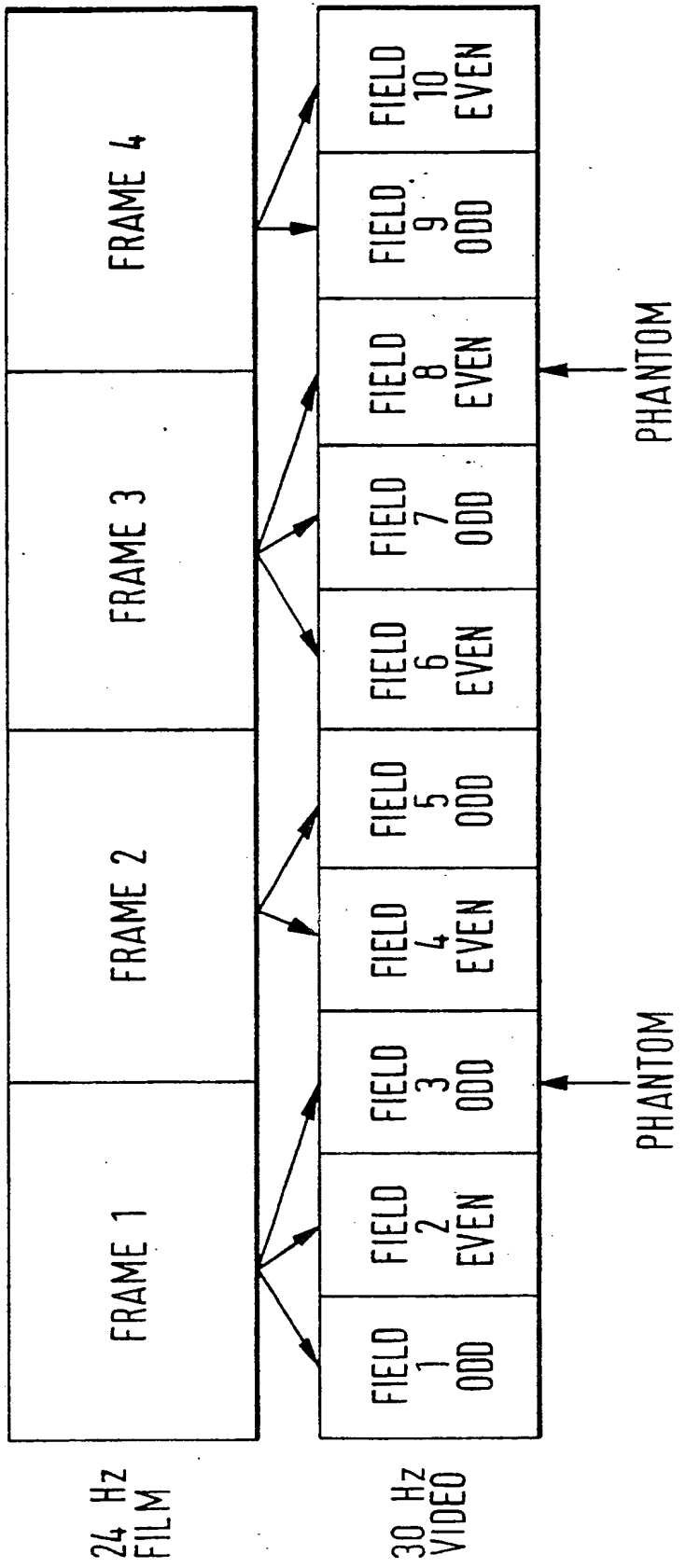
△ = PIXEL VECTOR A  
5 = VECTOR UNDER TEST

FIG. 48.



21/45

FIG. 49



22/45

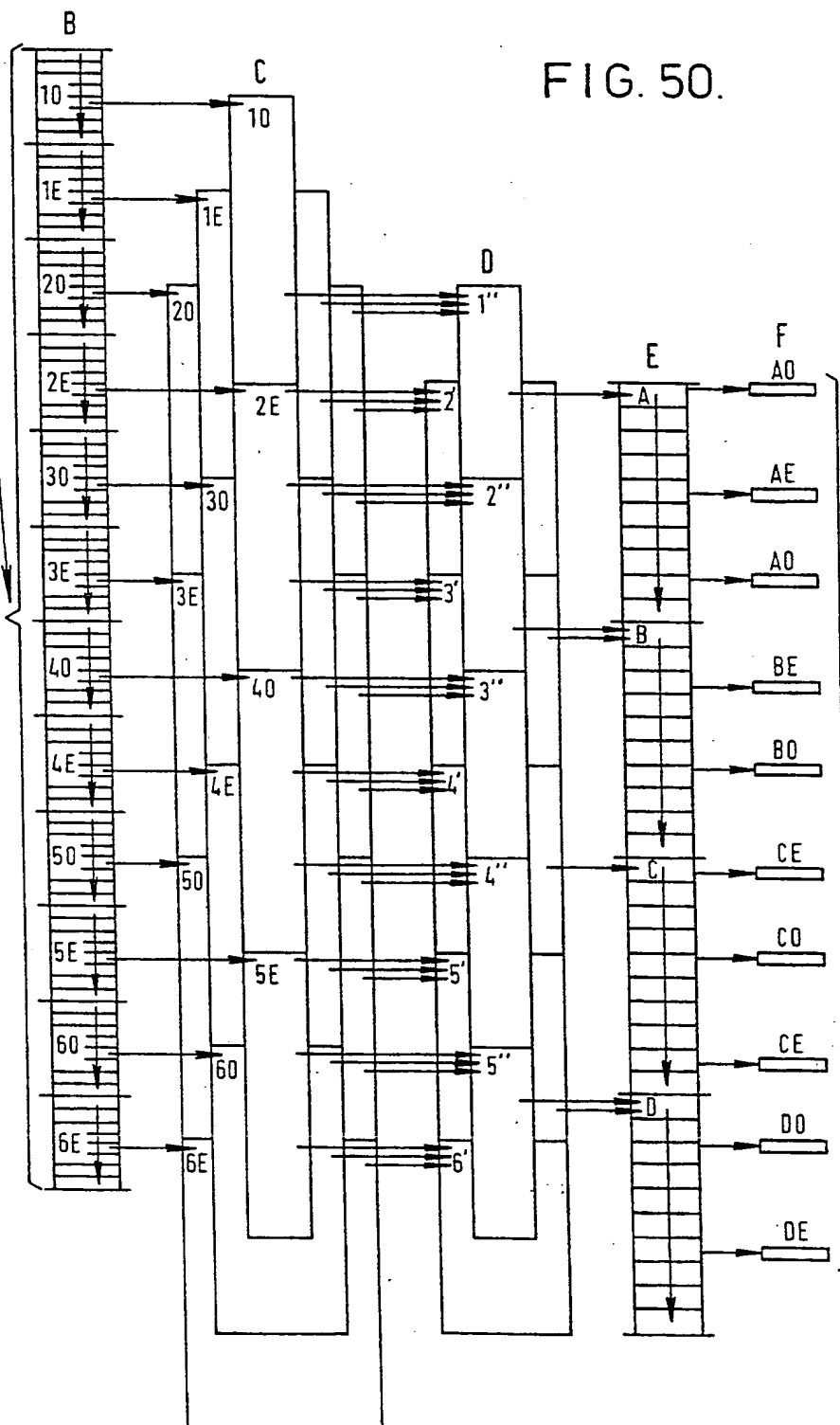
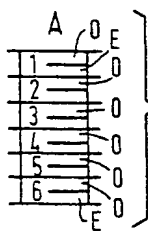
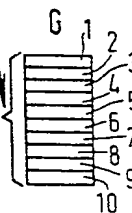


FIG. 50.



23/45

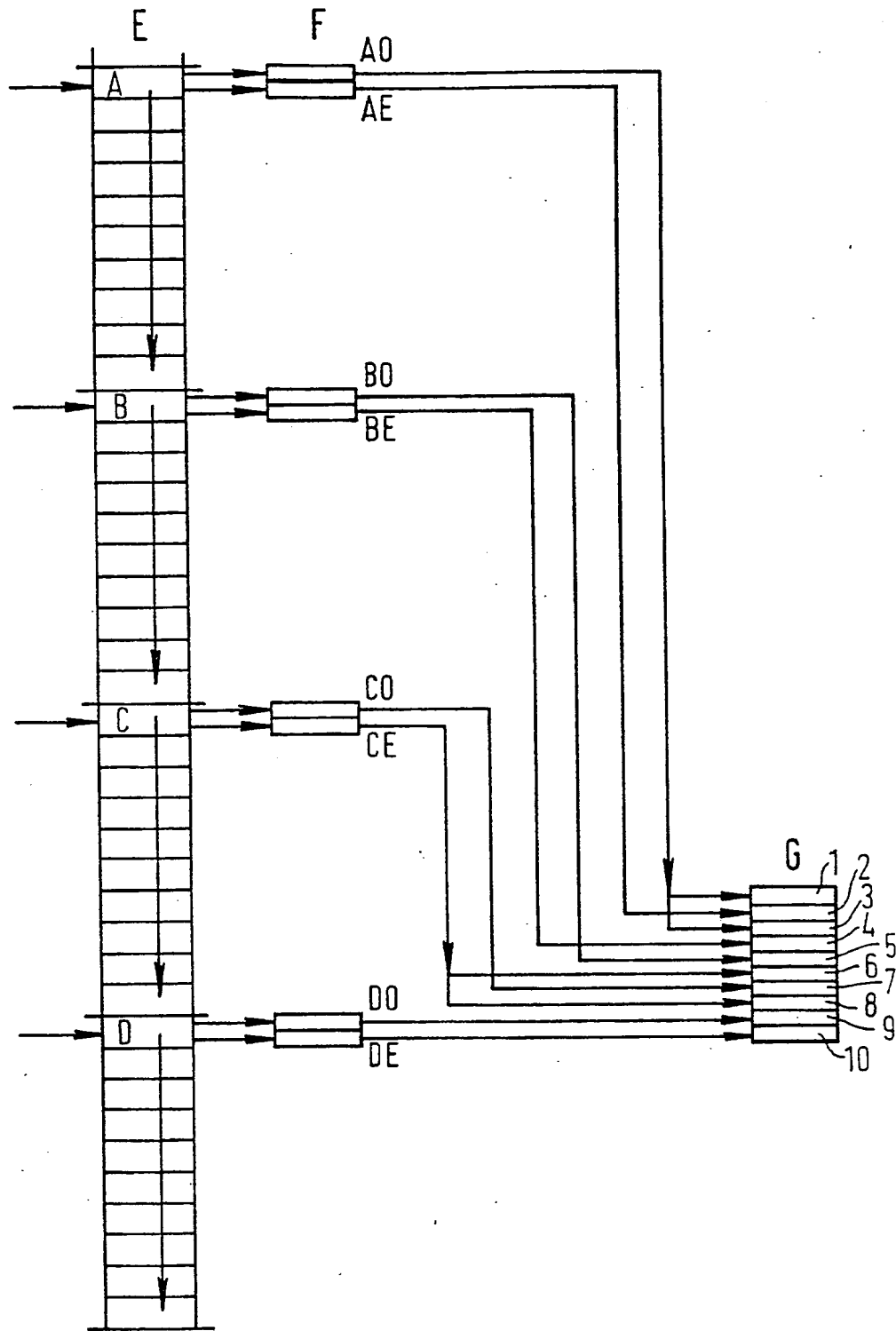
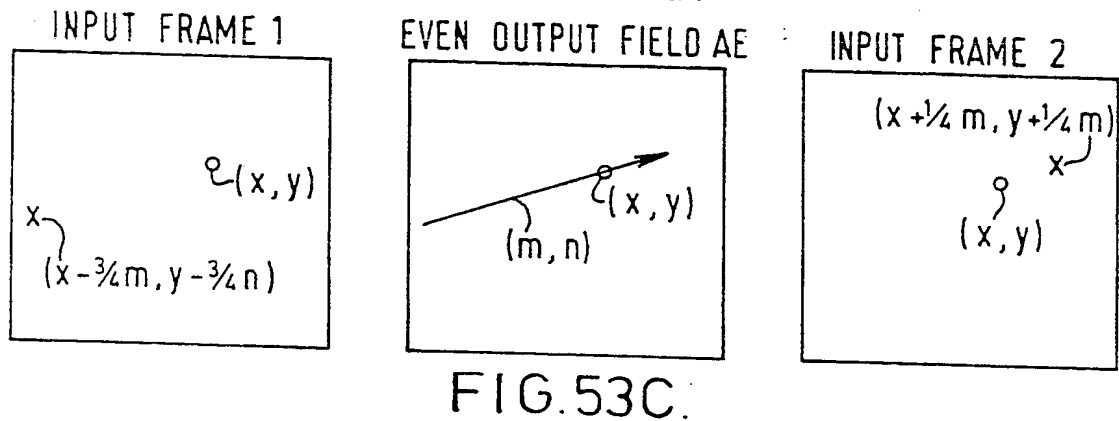
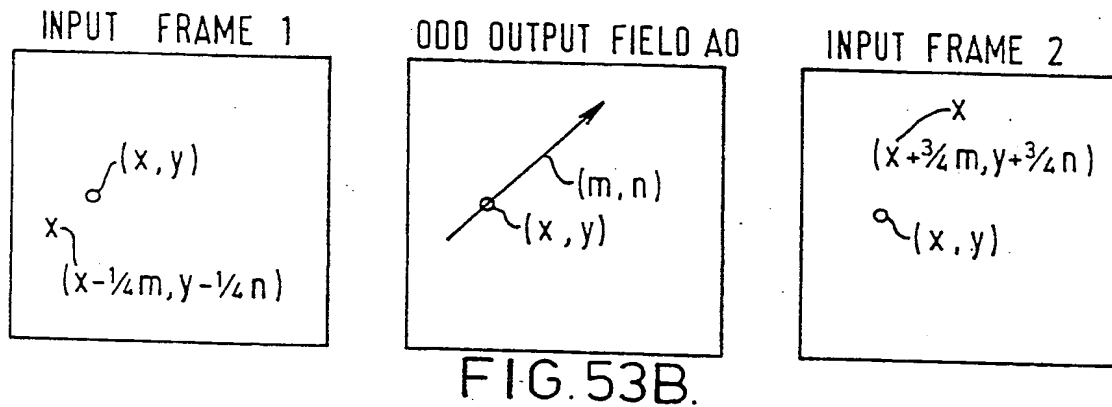
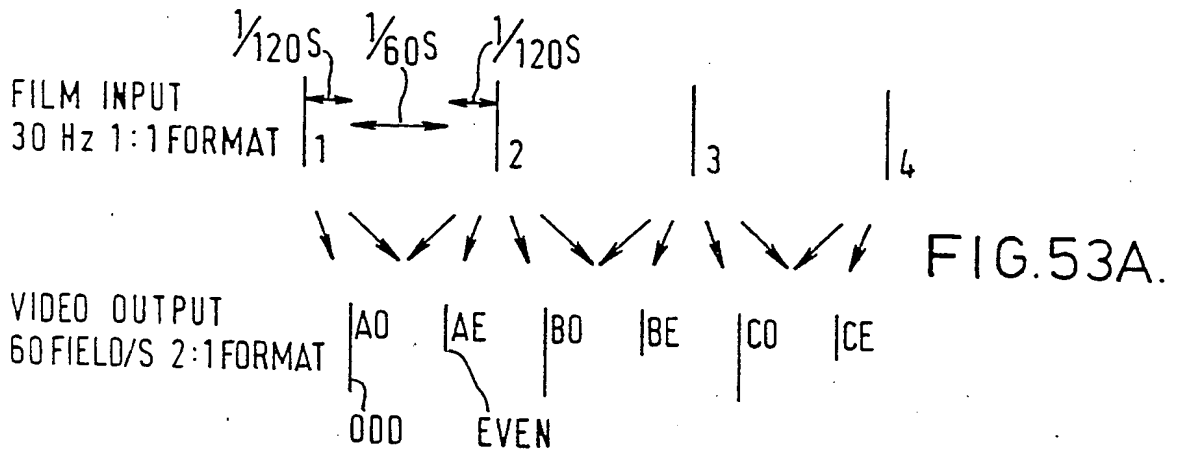
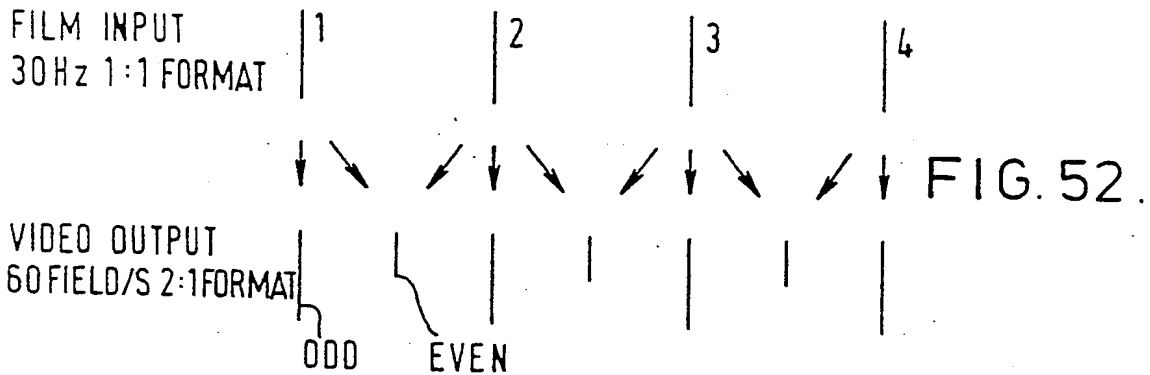


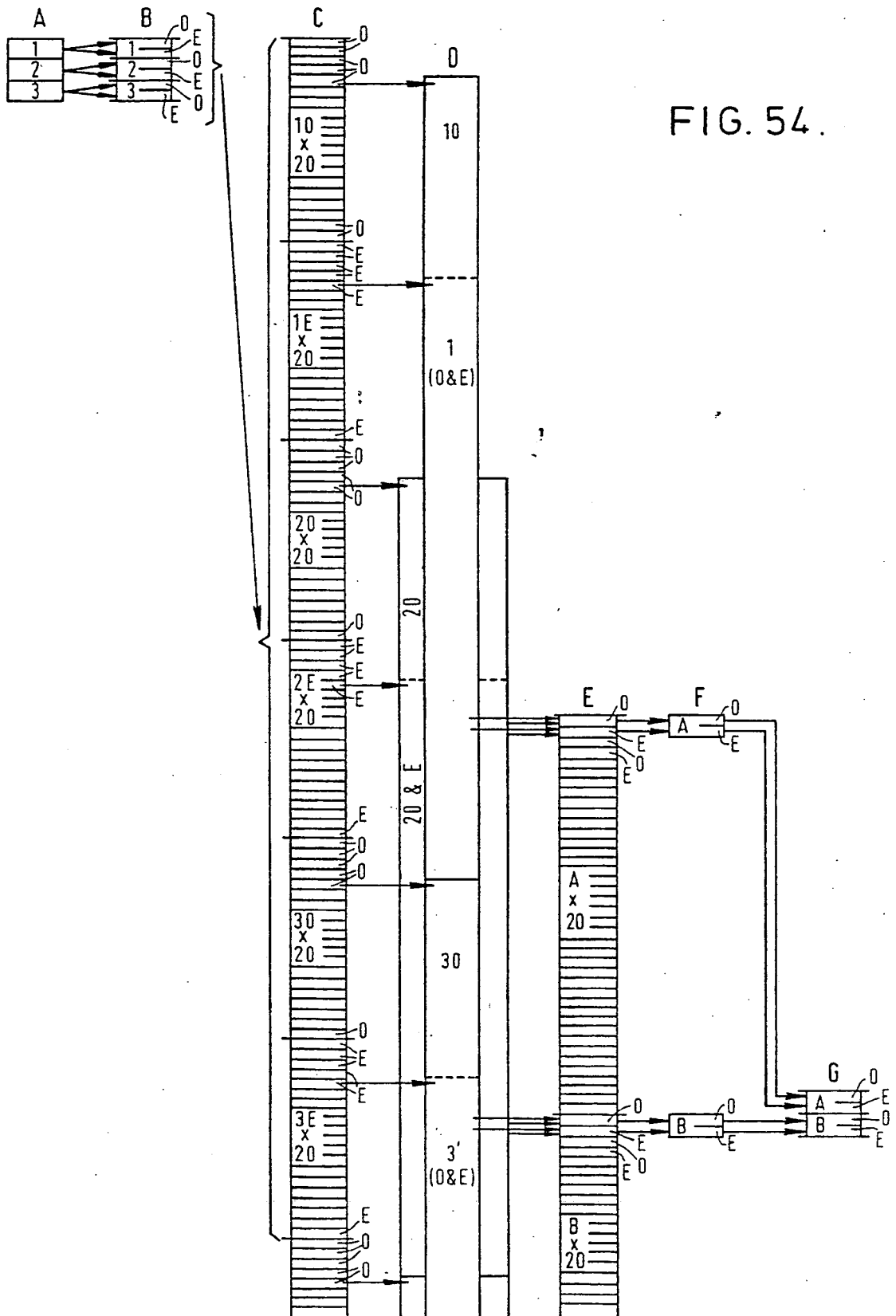
FIG. 51.



24/45



25/45



26/45

FILM INPUT  
24 Hz 1:1

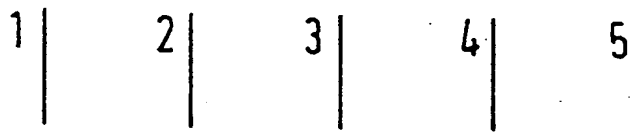
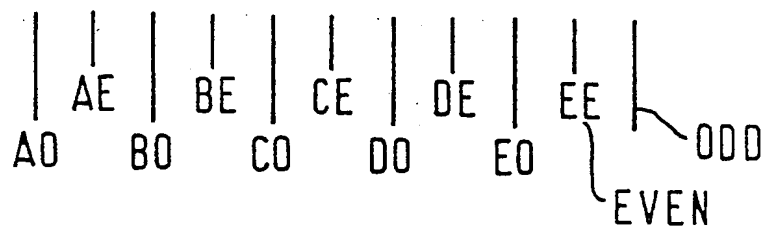


FIG. 55A.

VIDEO OUTPUT  
60 FIELD/S  
2:1 FORMAT



FILM INPUT  
24 Hz 1:1



FIG. 55B.

FILM OUTPUT  
30 Hz 1:1



27/45

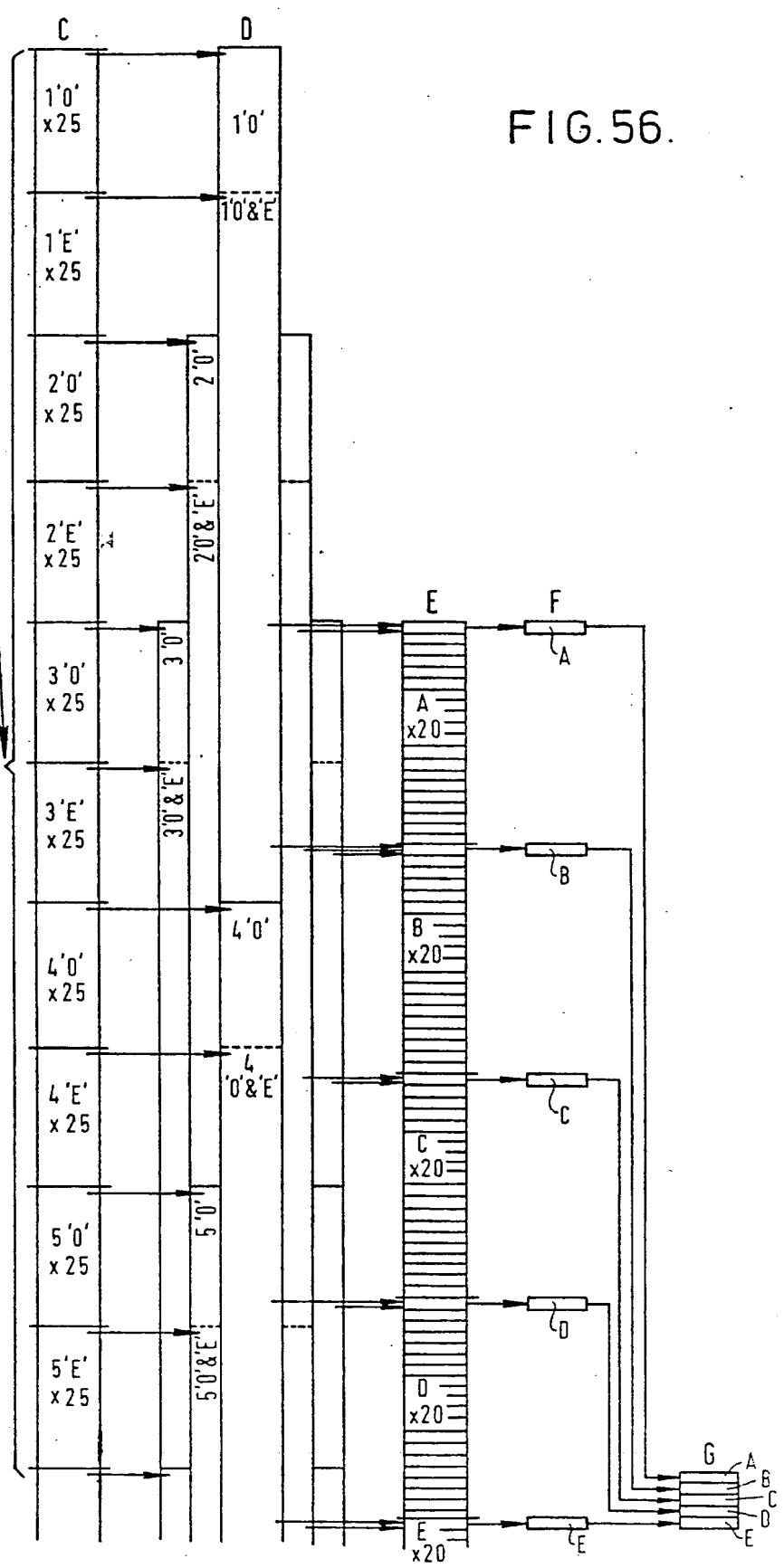
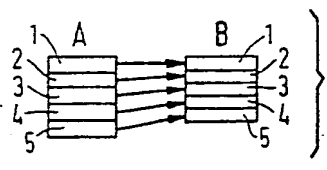


FIG. 56.

28/45

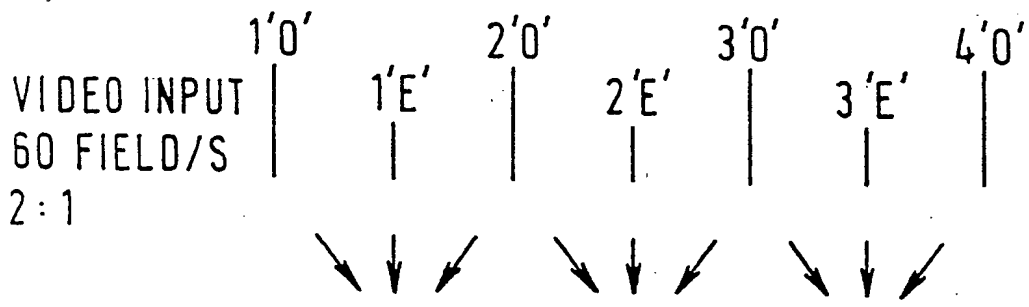


FIG. 57.

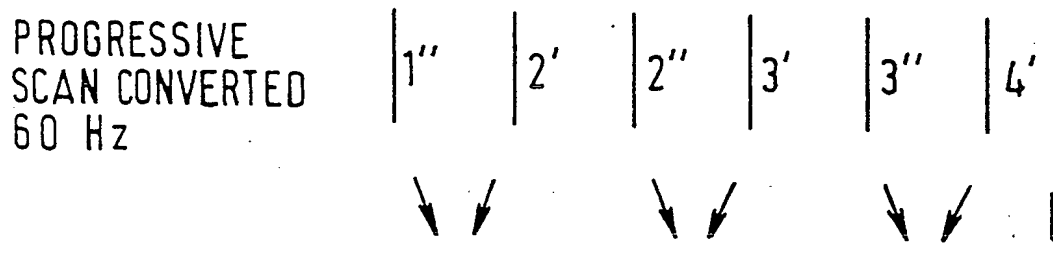
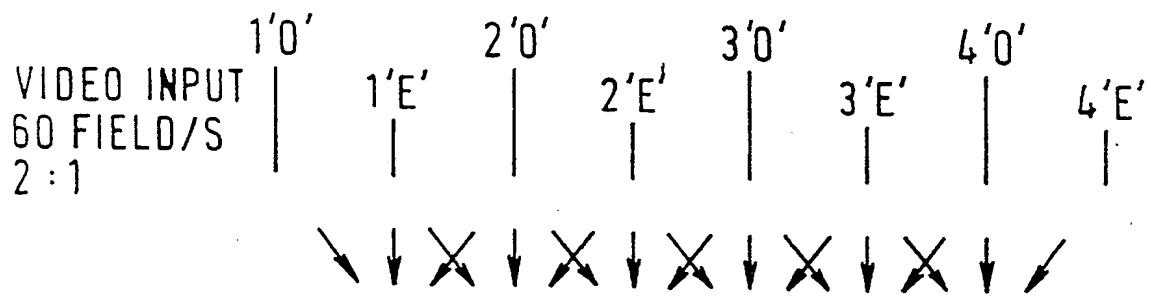
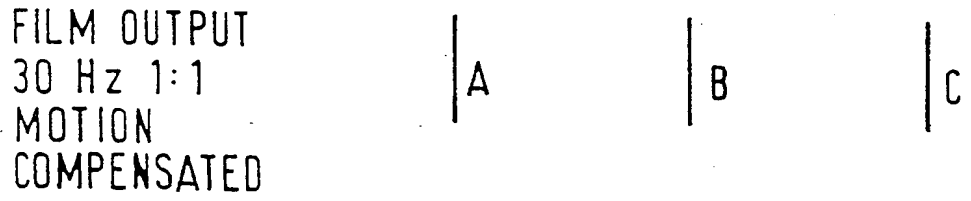
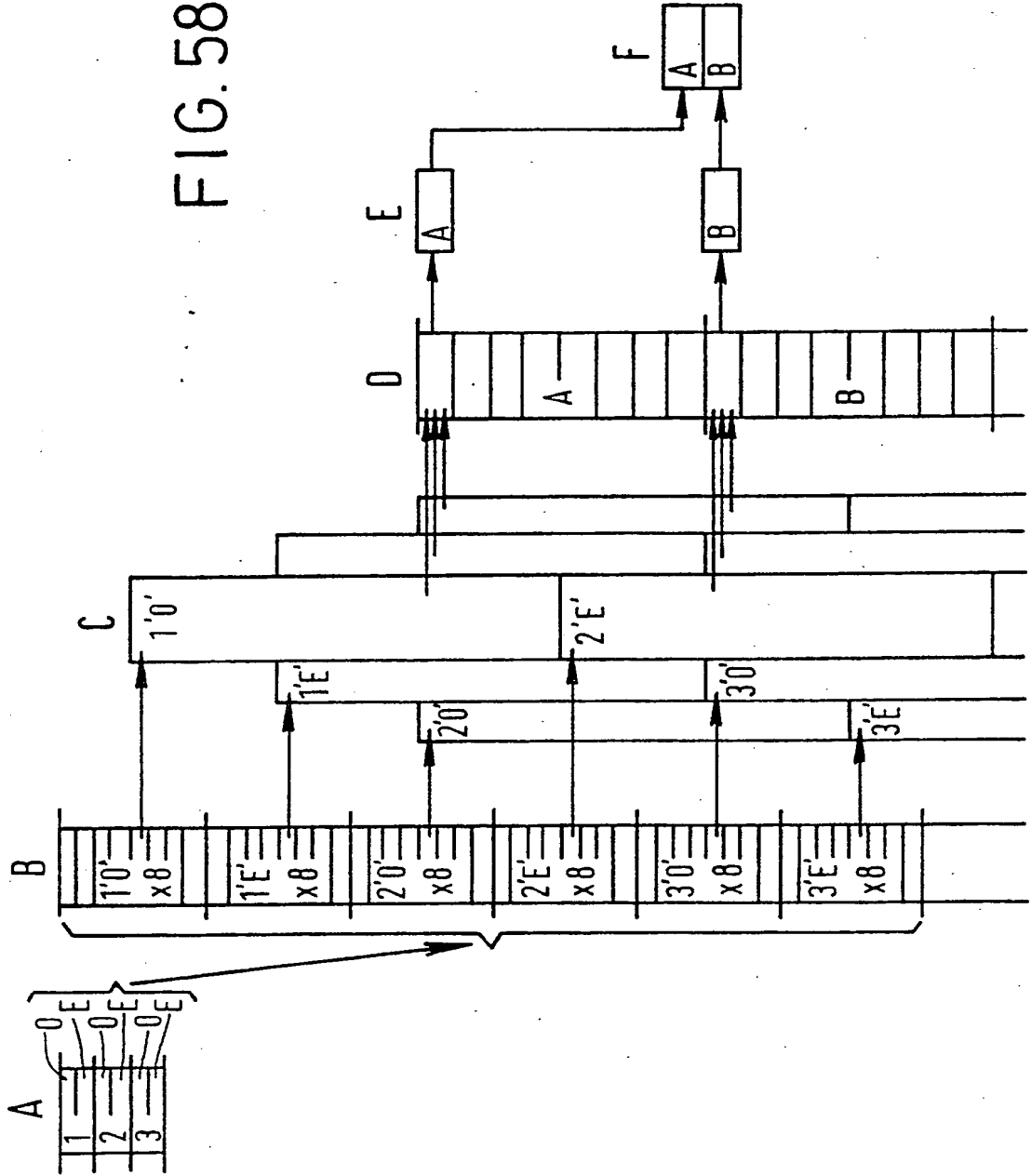


FIG. 59.



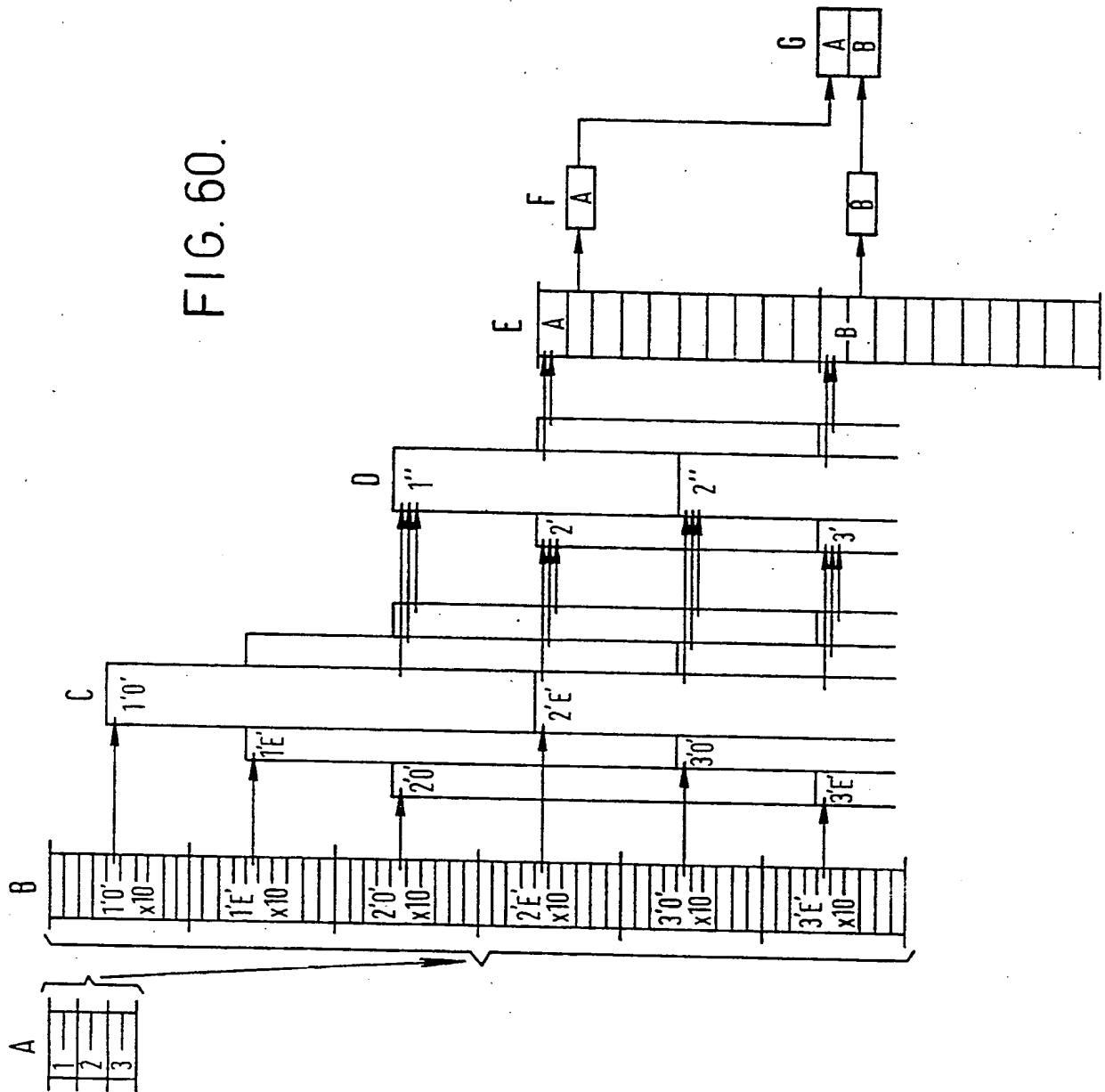
29/1/5

FIG. 58.

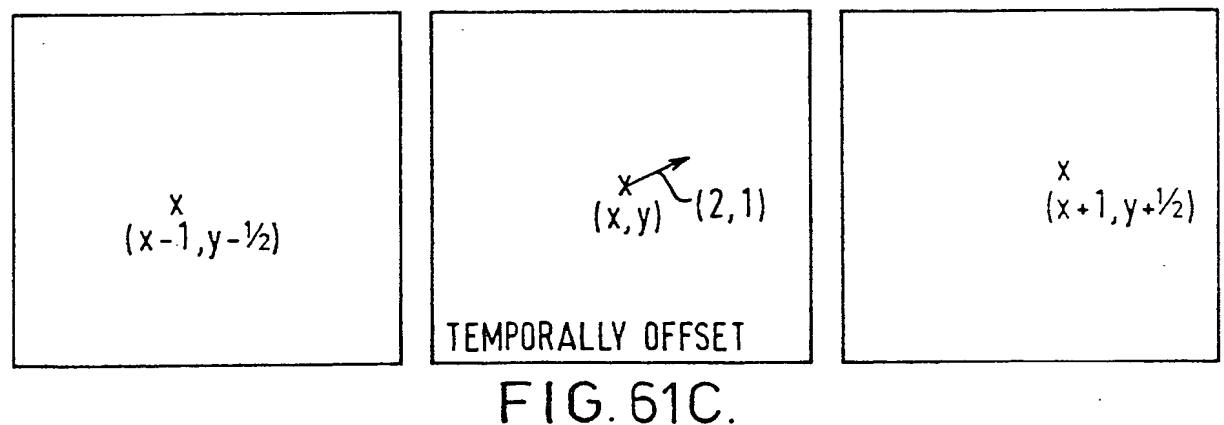
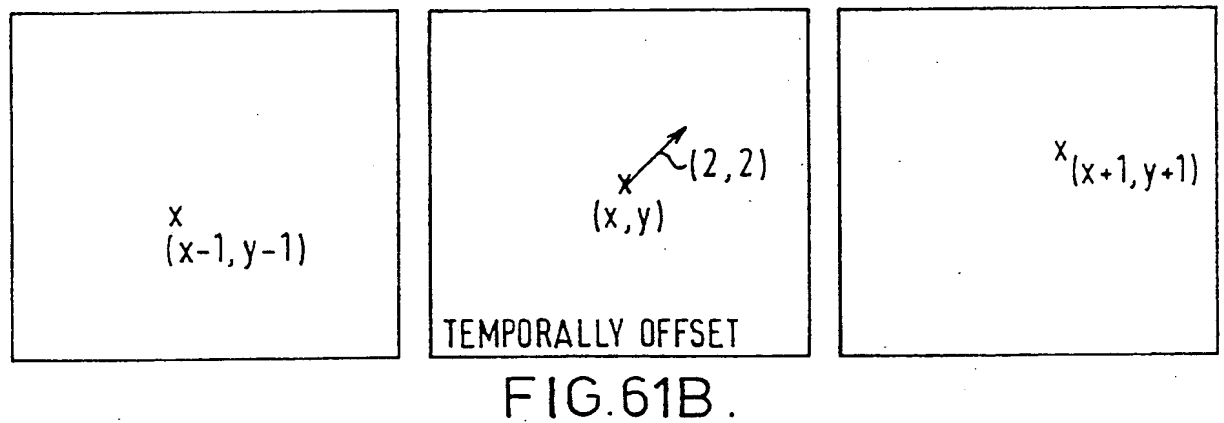
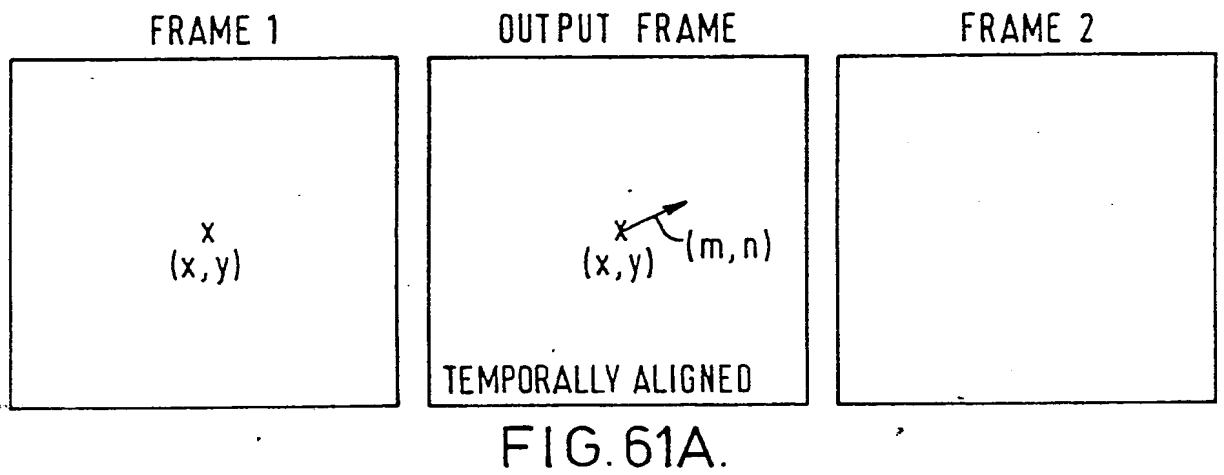


30/45

FIG. 60.

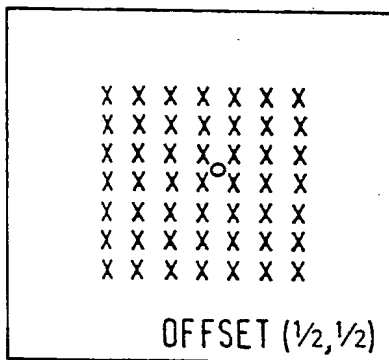
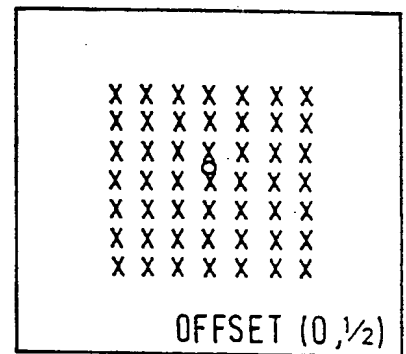
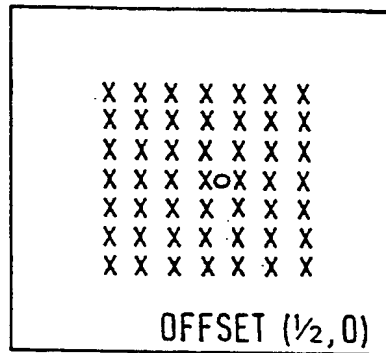
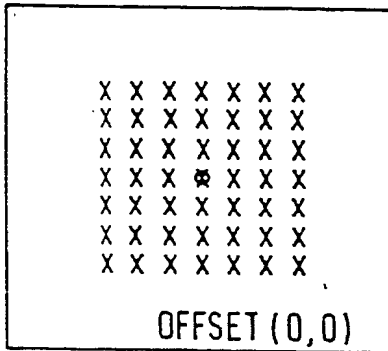


31/45





32/45



33/45

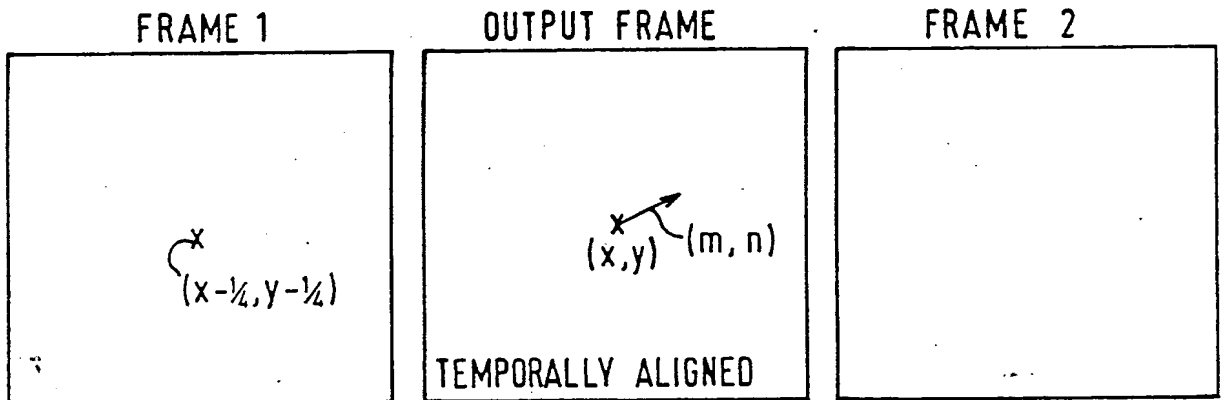


FIG. 63A.

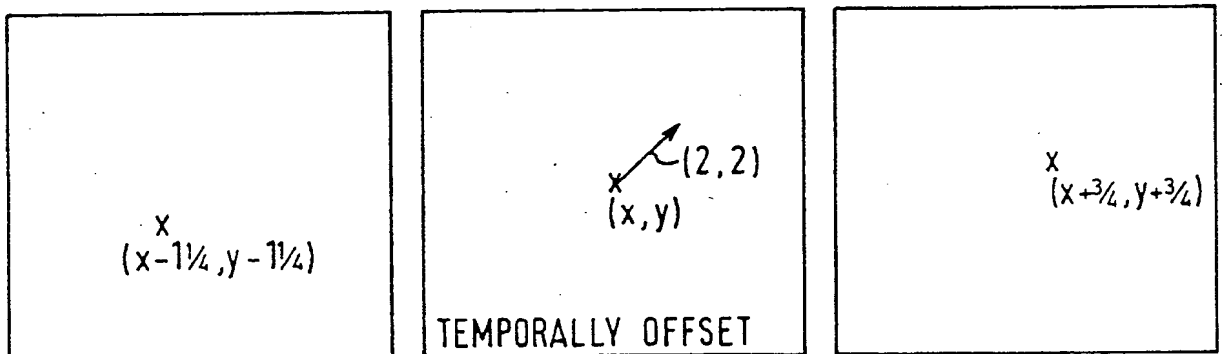


FIG. 63B.

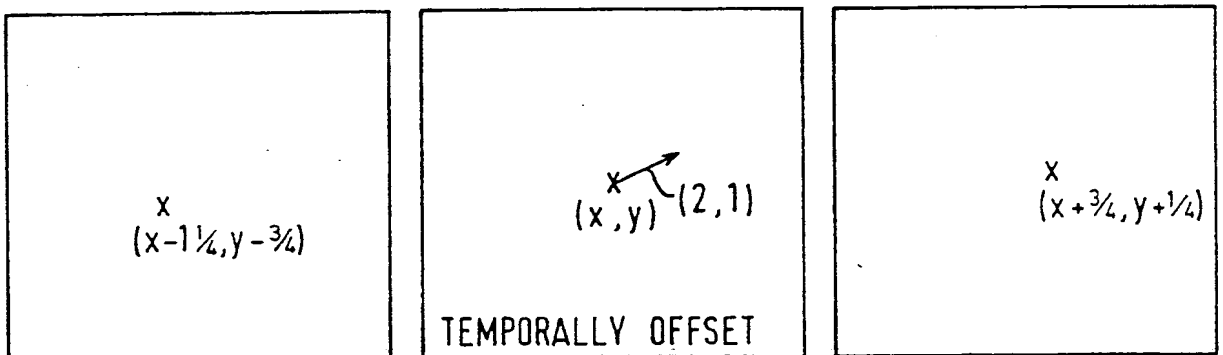


FIG. 63C.

34/45

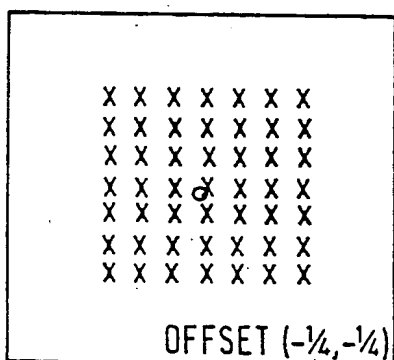


FIG. 64A.

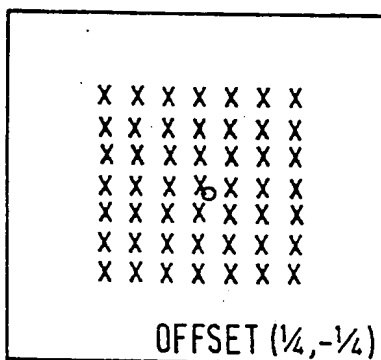


FIG. 64B.

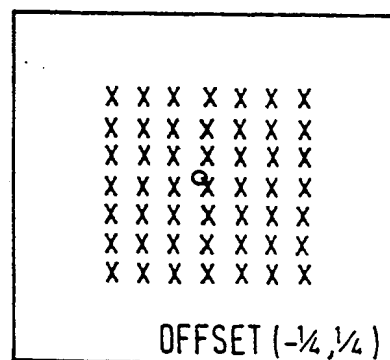


FIG. 64C.

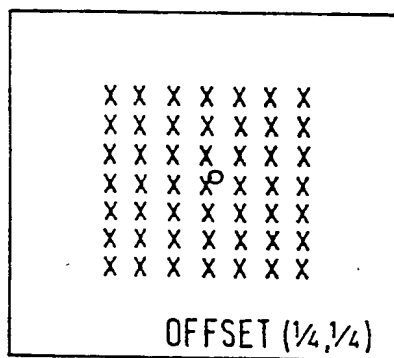


FIG. 64D.

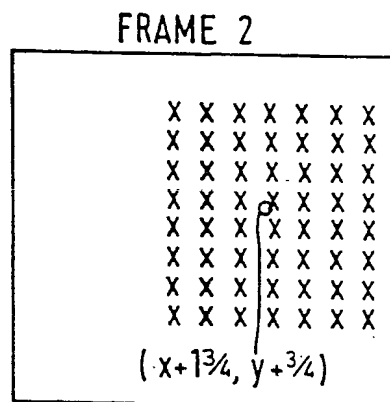
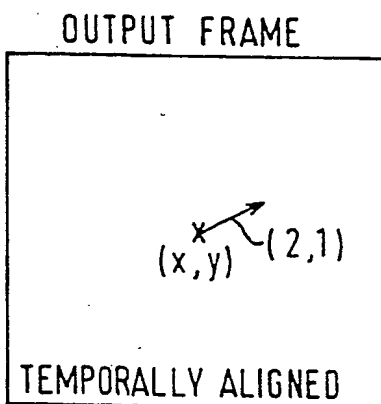
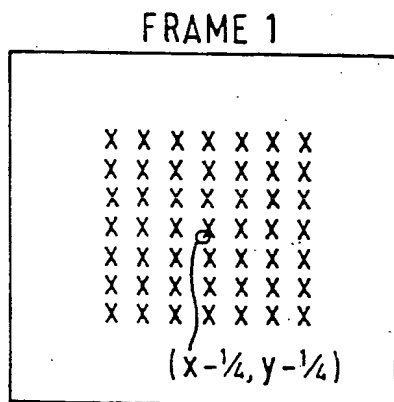


FIG. 65.

35/45

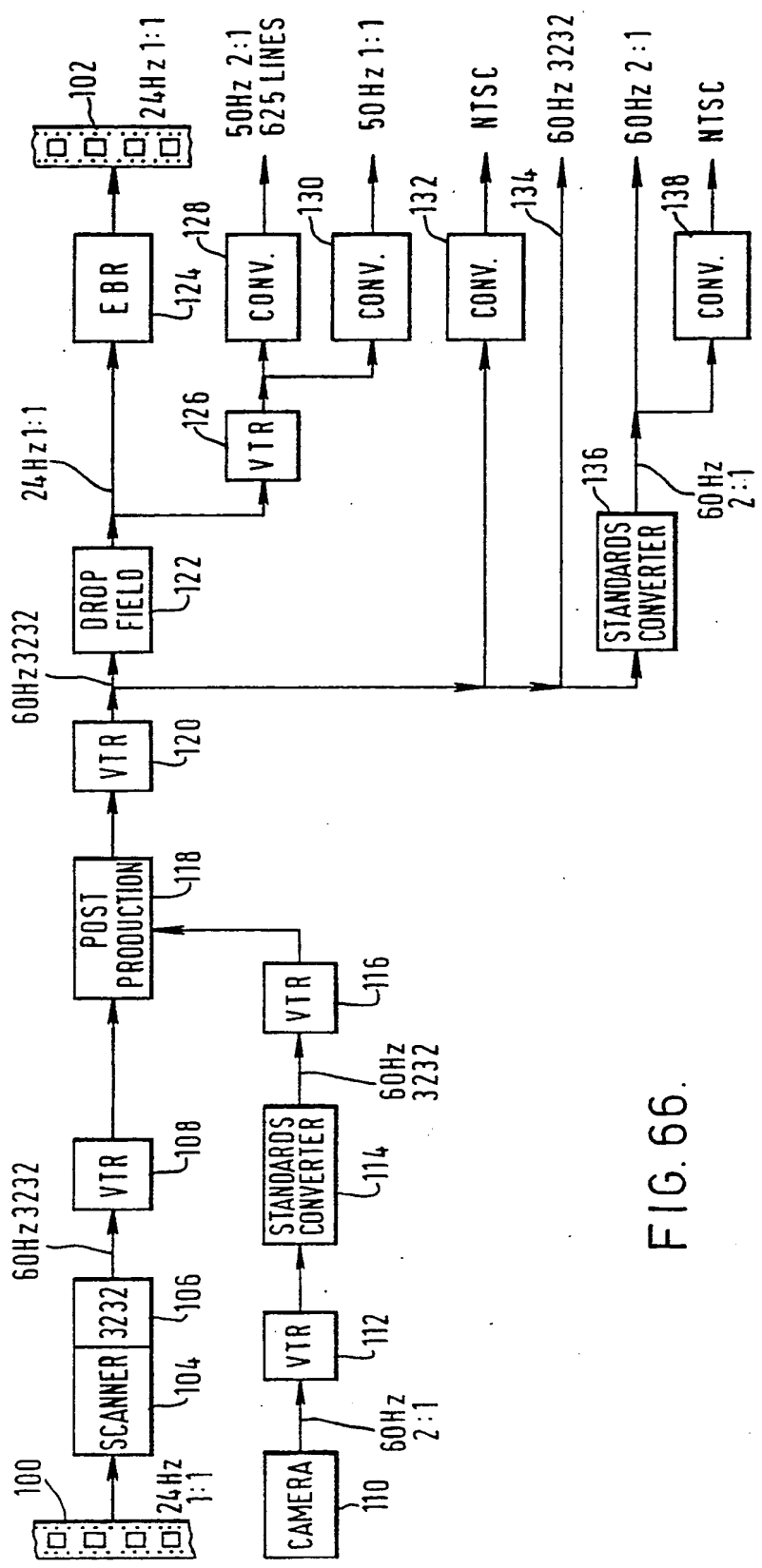


FIG. 66.

36/45

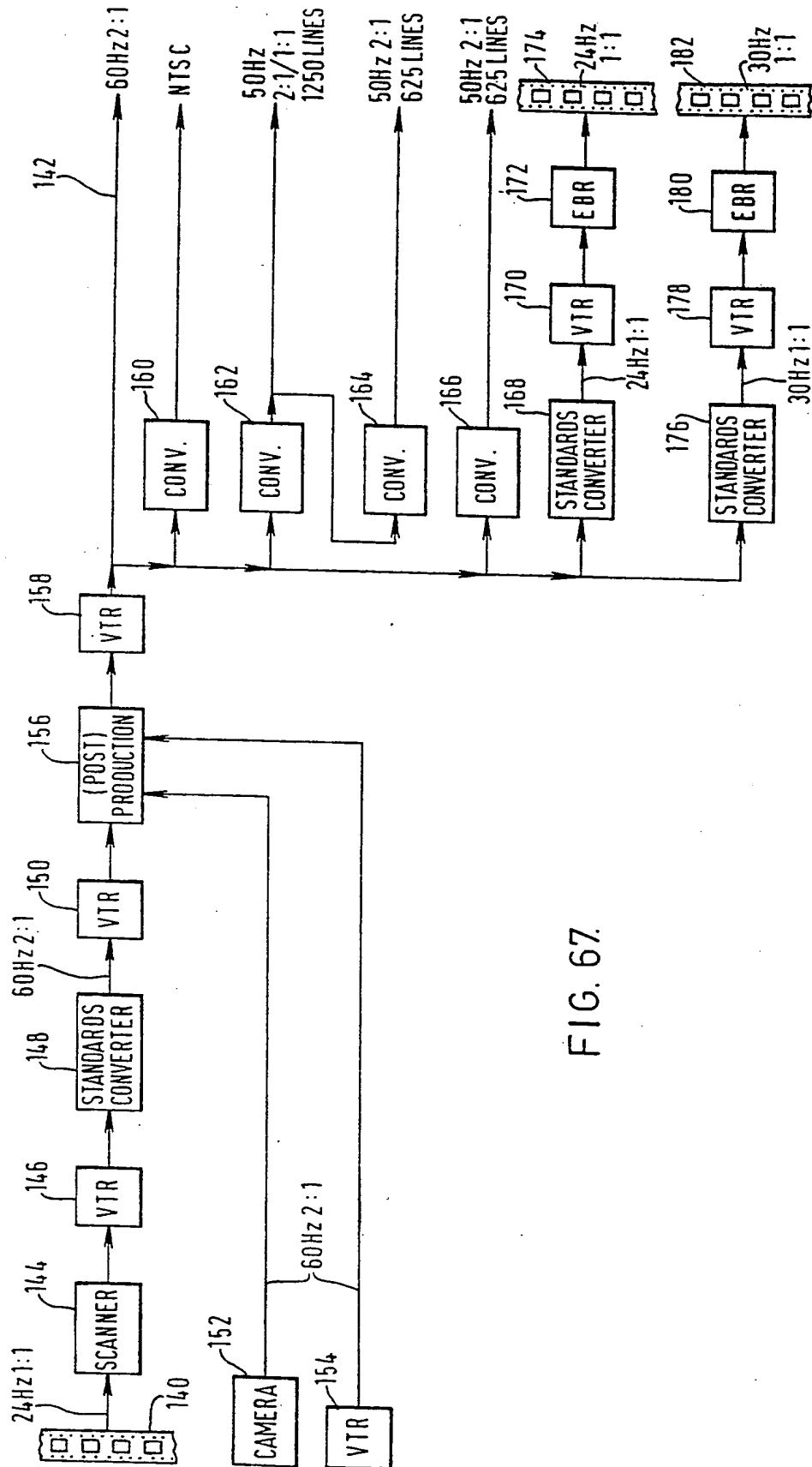


FIG. 67.

The diagram illustrates a video production system with the following components and signal paths:

- Inputs:**
  - 30Hz 1:1:** A signal source feeding into a **SCANNER (194)** and a **CAMERA (198)**.
  - 60Hz 2:1:** A signal source feeding into a **CAMERA (202)** and a **STANDARDS CONVERTER (206)**.
- Processing and Storage:**
  - The **SCANNER (194)** outputs to a **VTR (196)**, which then feeds into **POST PRODUCTION (208)**.
  - The **CAMERA (198)** outputs to **POST PRODUCTION (208)**.
  - The **CAMERA (202)** outputs to a **VTR (200)**, which feeds into **POST PRODUCTION (208)**.
  - The **STANDARDS CONVERTER (206)** outputs to **POST PRODUCTION (208)**.
- Output Paths:**
  - POST PRODUCTION (208)** outputs to a **VTR (210)**.
  - The **VTR (210)** outputs to a **STANDARDS CONVERTER (214)**, which then feeds into a **VTR (216)**.
  - The **VTR (216)** outputs to a **CONV. (220)**, which produces **NTSC** output.
  - The **VTR (216)** also outputs to a **CONV. (222)**, which produces **50Hz 2:1/1:1 1250 LINES** output.
  - The **VTR (216)** outputs to a **CONV. (224)**, which produces **50Hz 2:1 625 LINES** output.
  - The **VTR (216)** outputs to a **STANDARDS CONVERTER (226)**, which feeds into a **VTR (228)**.
  - The **VTR (228)** outputs to an **EBR (230)**.
  - The **EBR (230)** outputs to a **CONV. (234)**, which produces **24Hz 1:1** output.

FIG. 68.

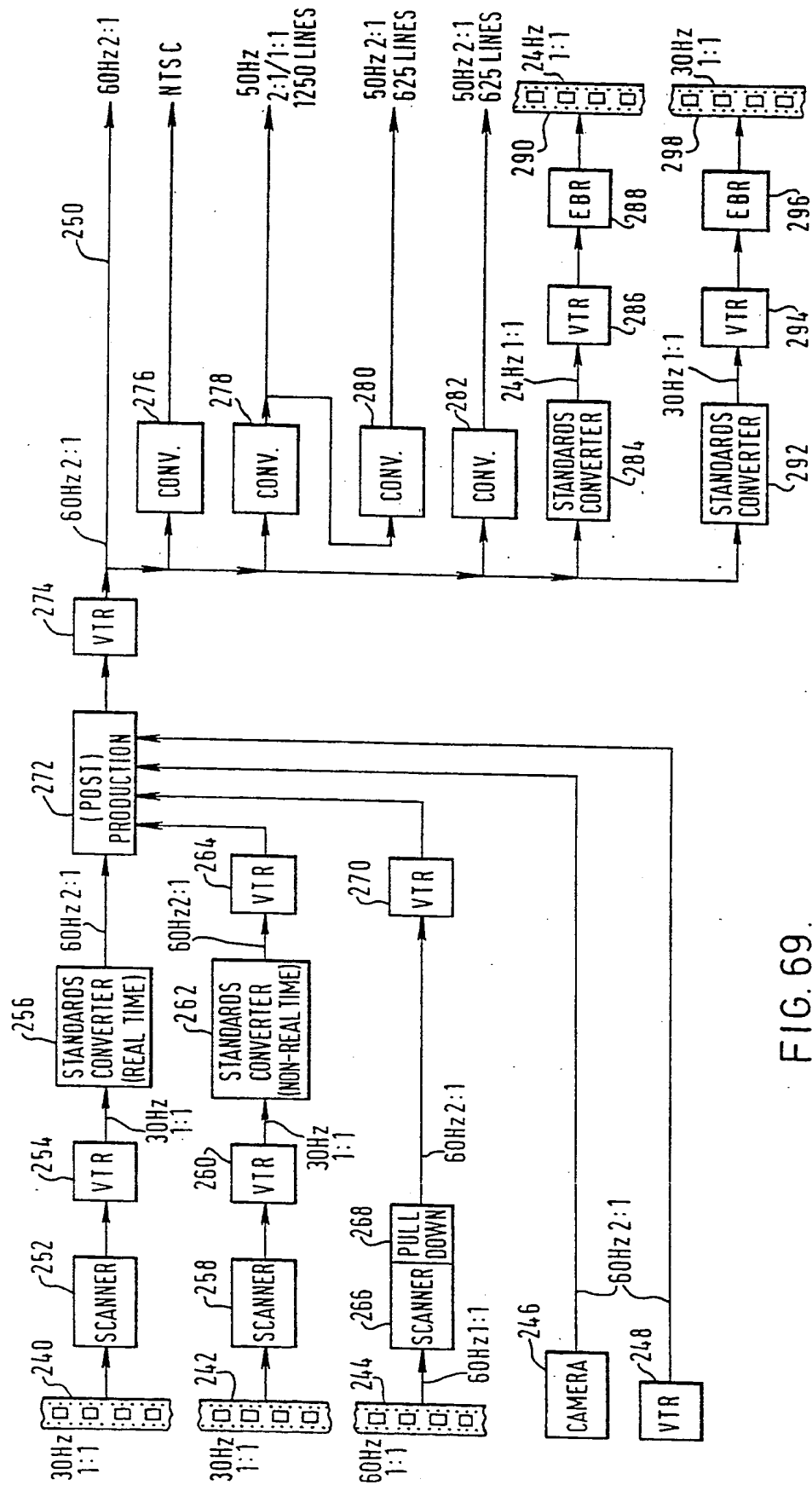


FIG. 69.

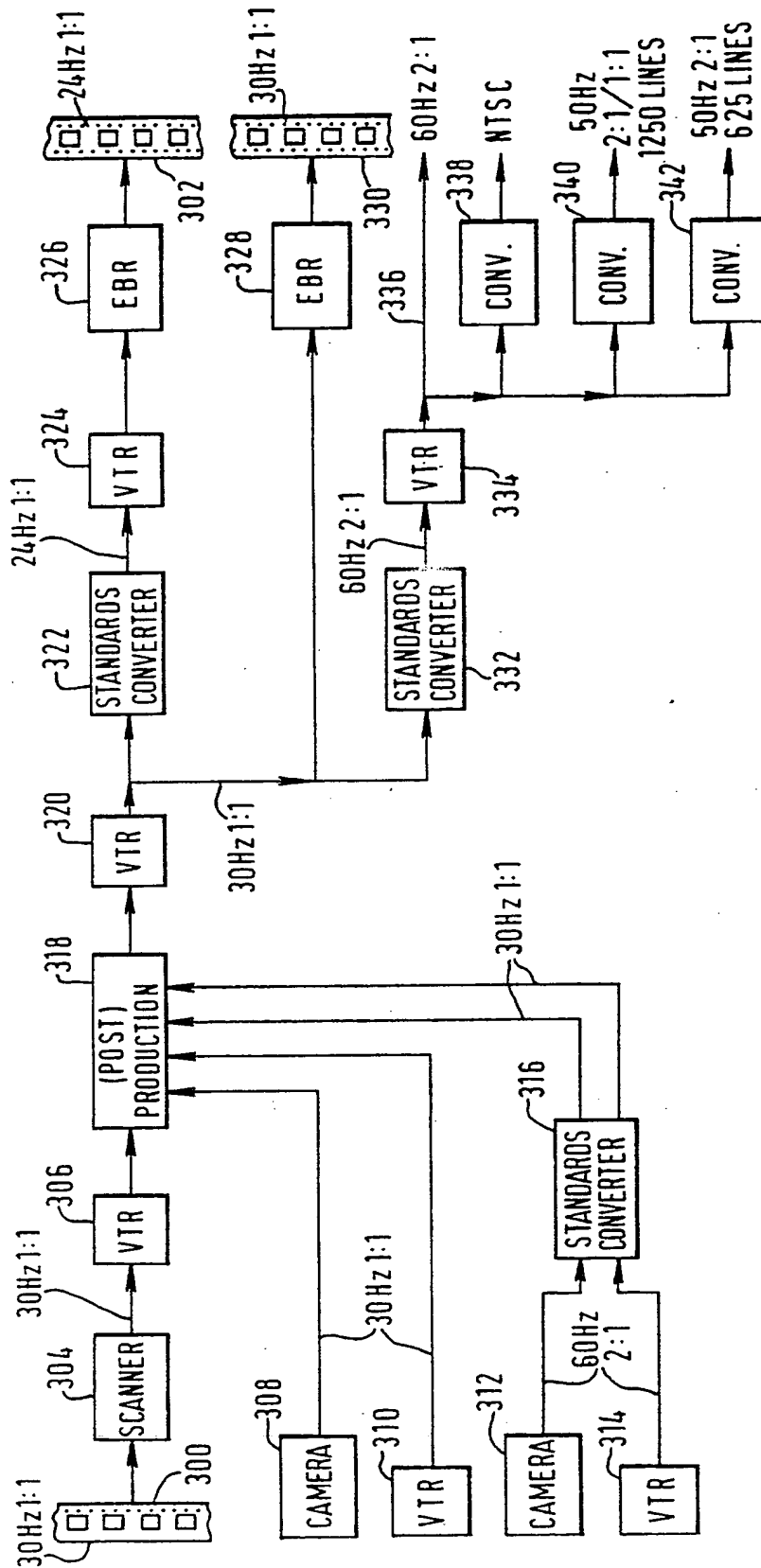


FIG. 70.



40/45

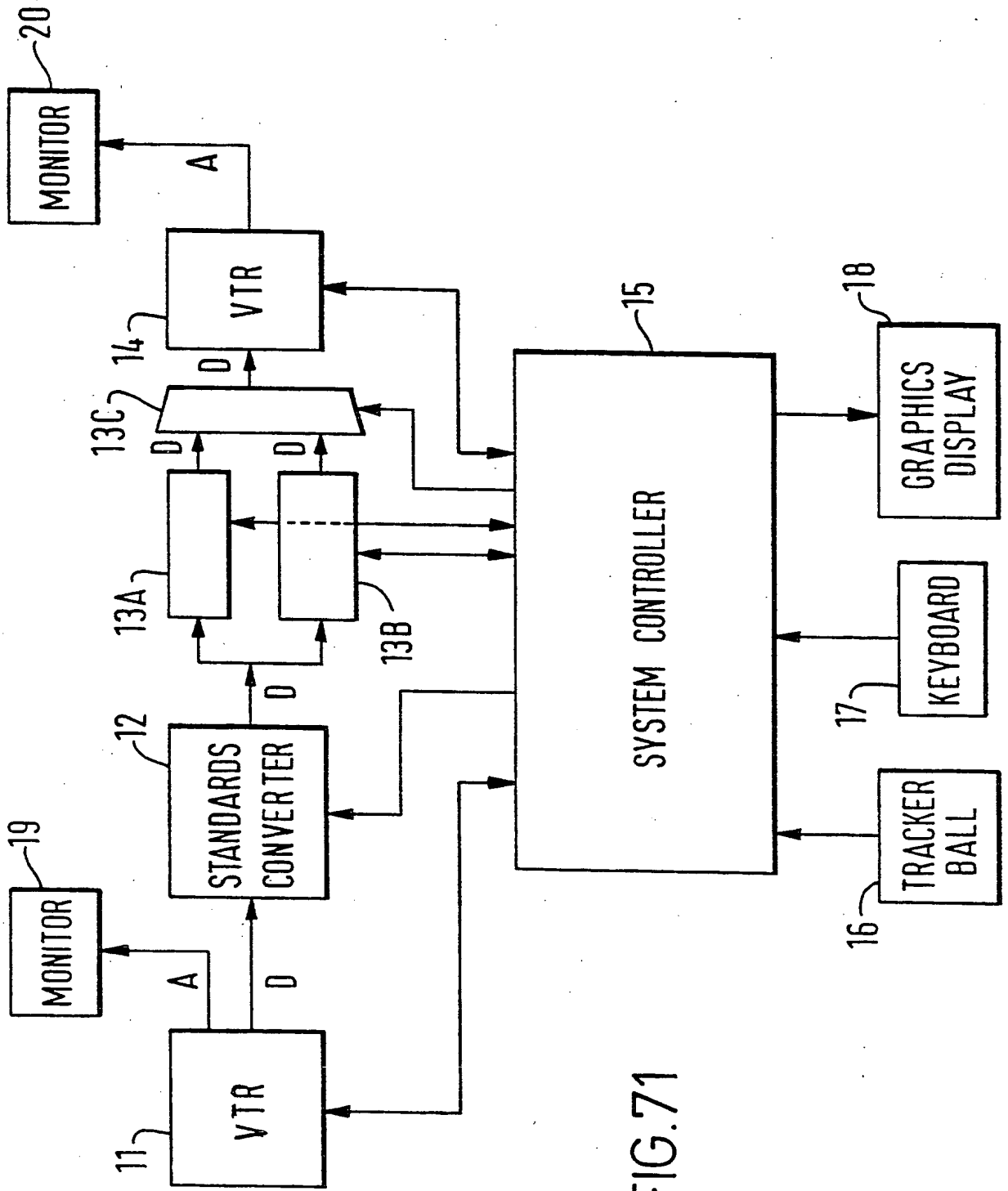


FIG. 71

41/45

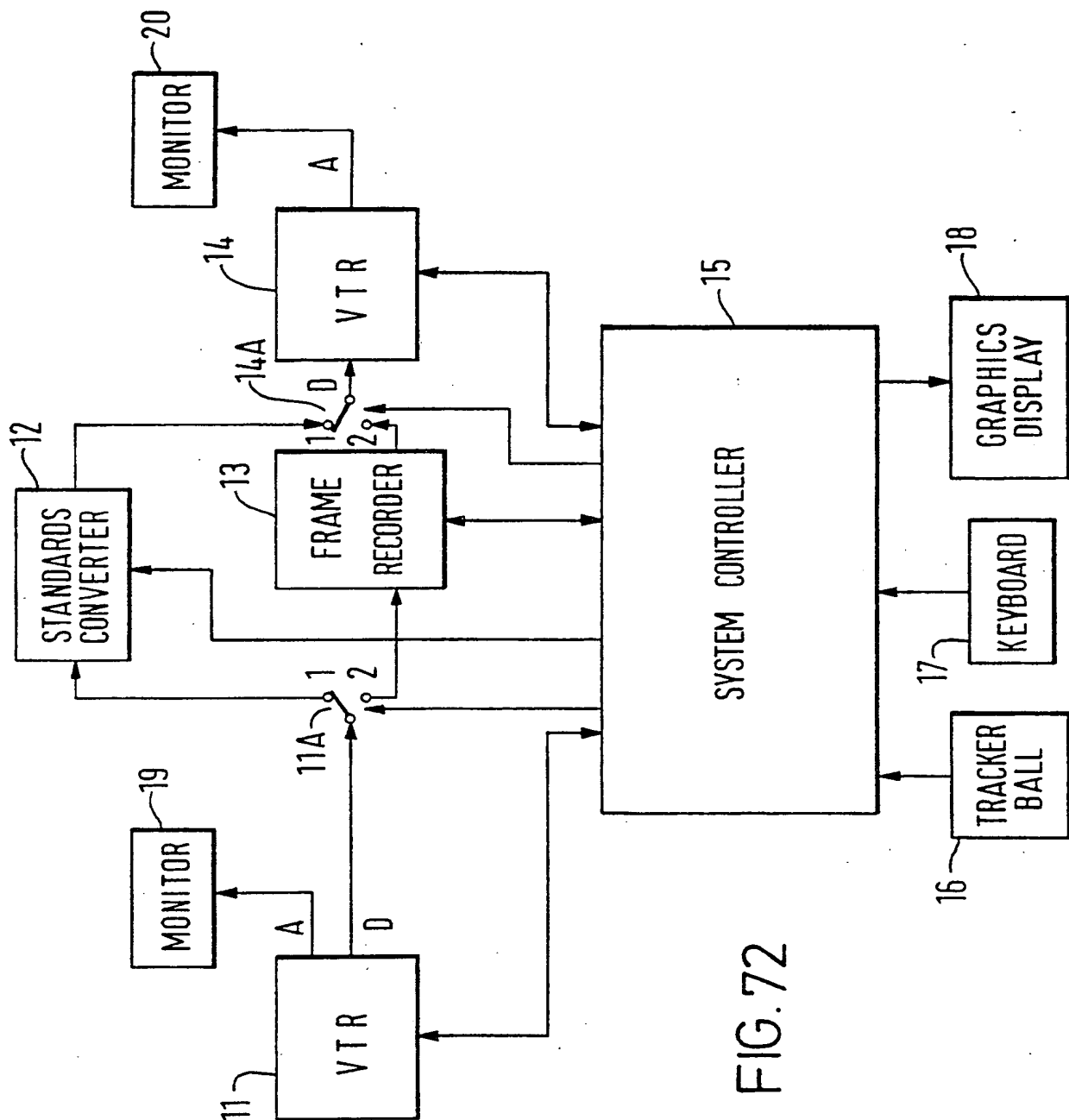


FIG. 72

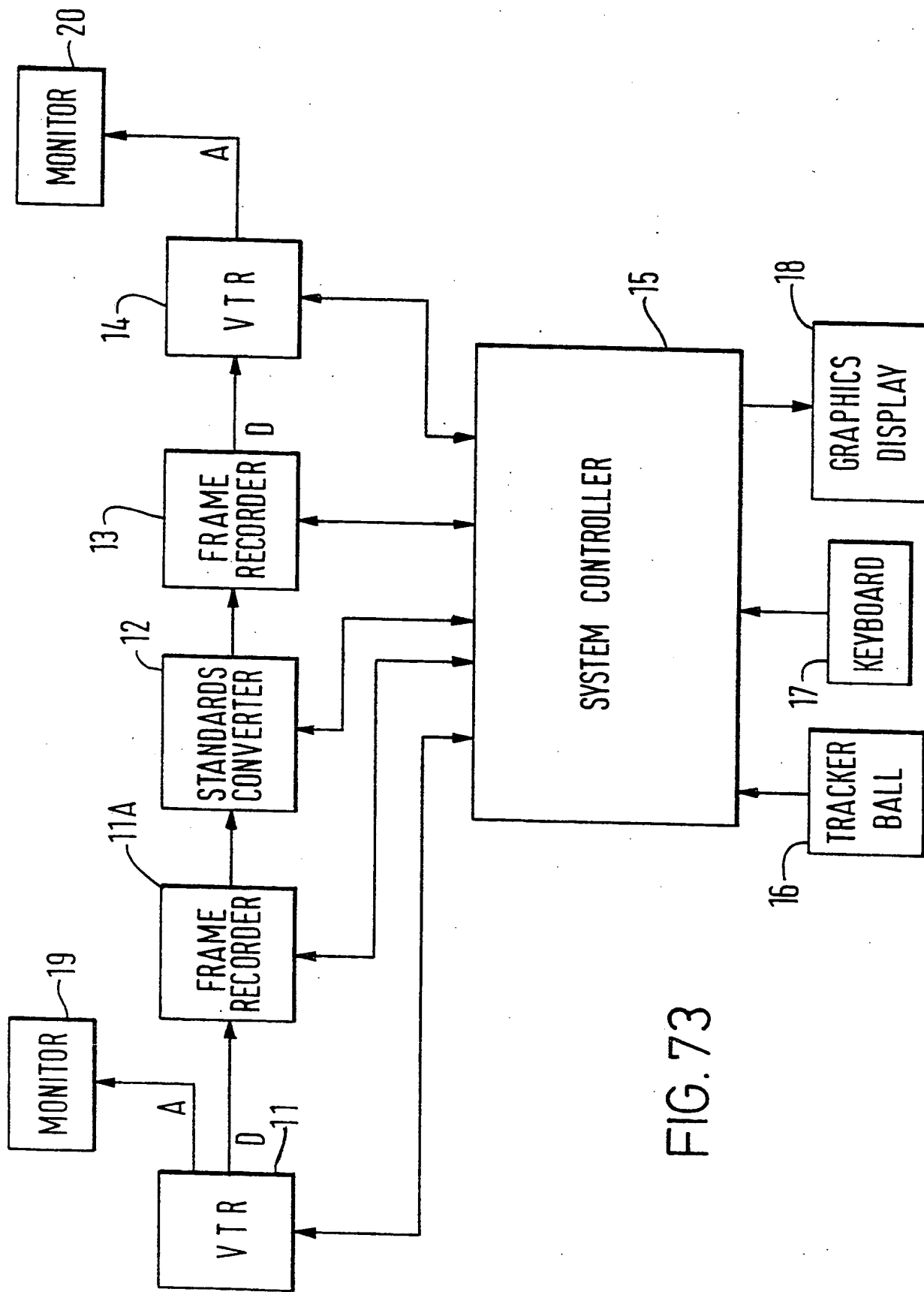


FIG. 73

43 / 45

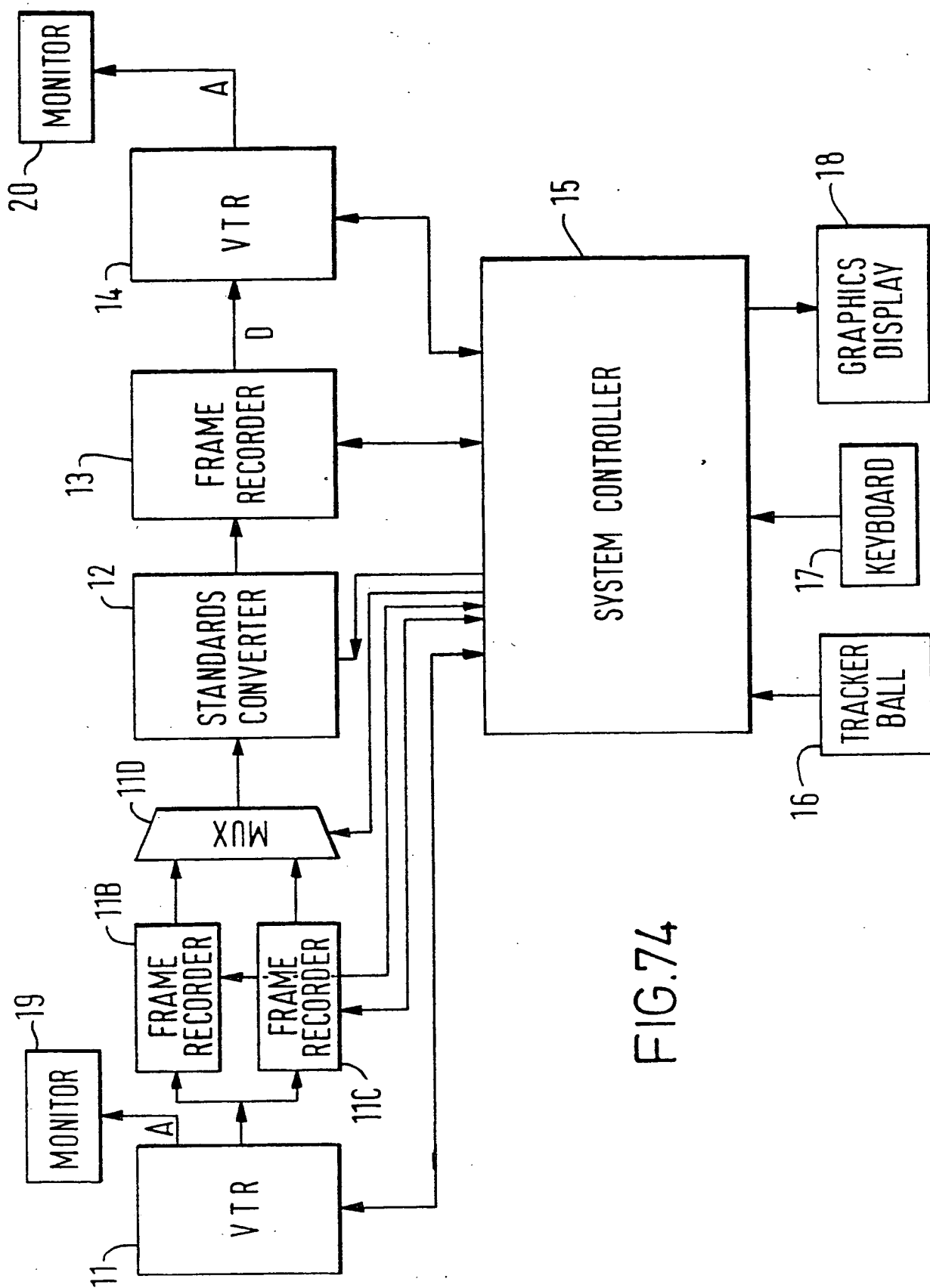


FIG. 74

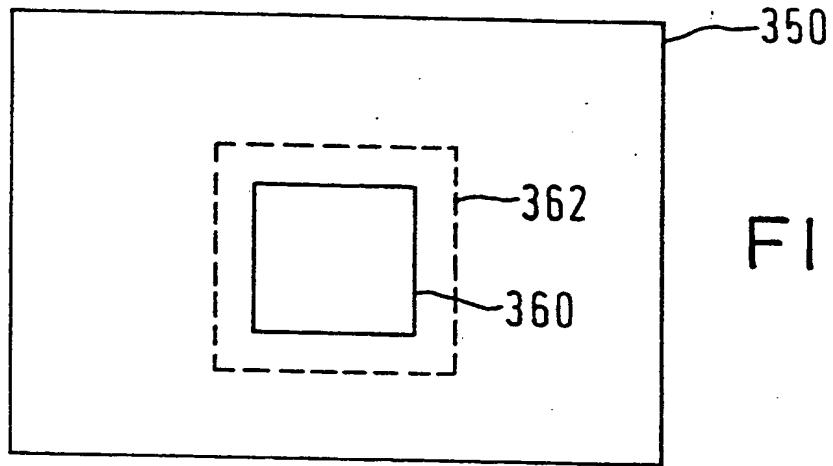


FIG. 75.

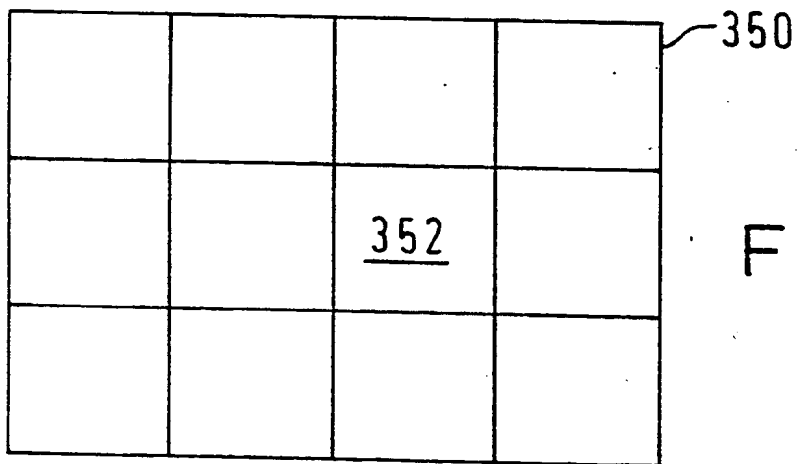


FIG. 76.

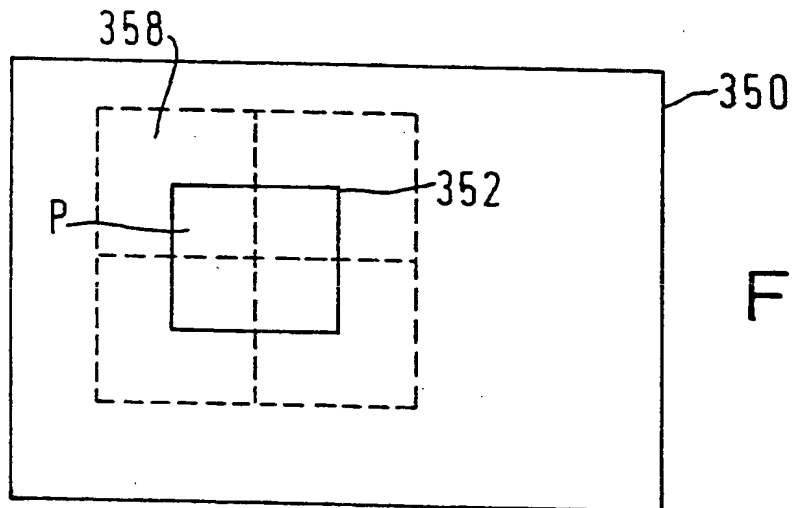


FIG. 77.

45/45

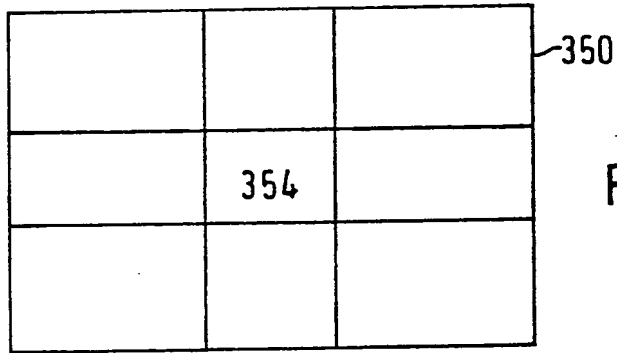


FIG. 78.

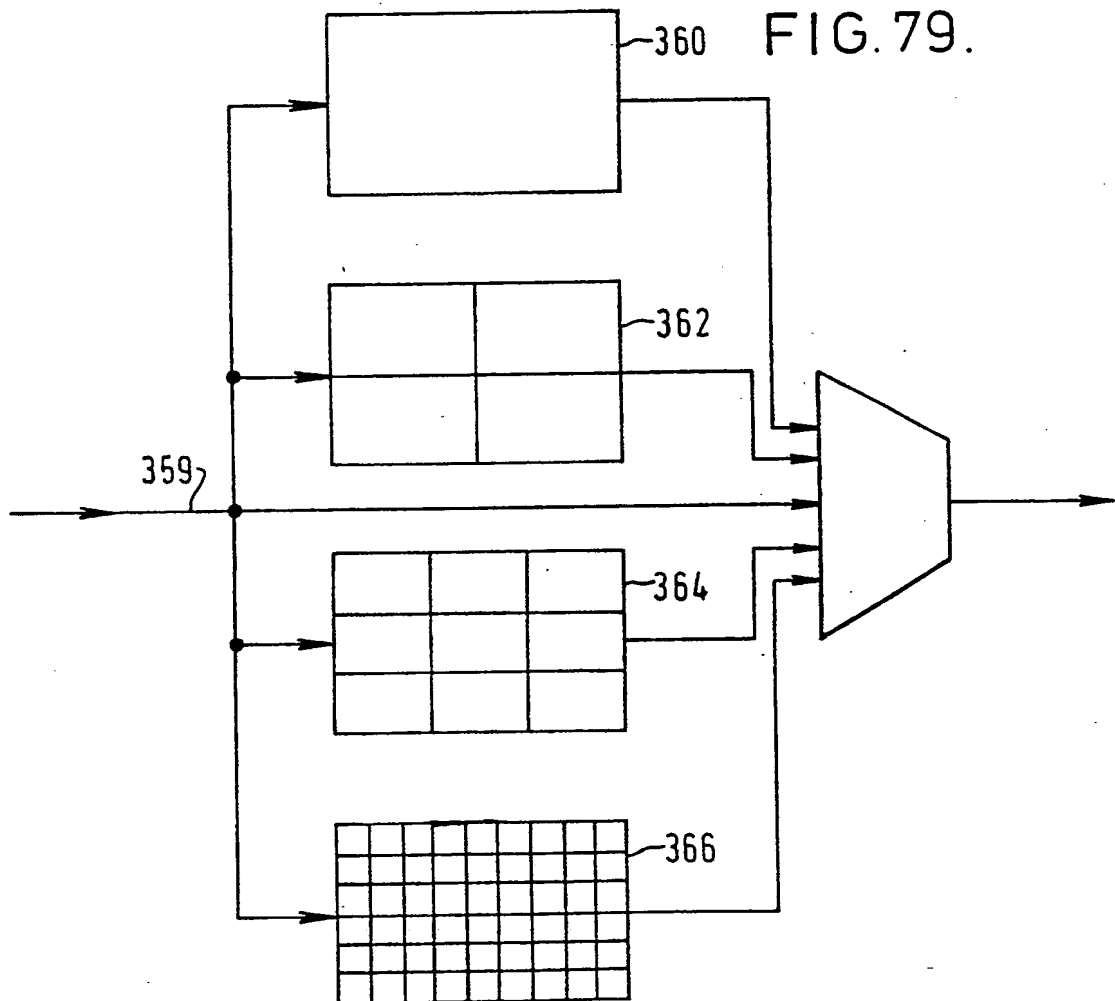


FIG. 79.

MOTION COMPENSATED INTERPOLATION OF IMAGES

This invention relates to the formation of an output image by motion compensated interpolation between a pair of input images.

5        Patent Application GB 8909656.4 describes an arrangement for converting 60 field/s 2:1 interlace HDVS to 24 Hz 1:1 film format. In that arrangement, the output frames are either temporally aligned with respective input frames or temporally offset by one half of an input field period ( $1/120$ s). In the case of temporal alignment, an output  
10 frame is based upon a respective progressive format frame developed from an input field and temporally adjacent input fields, whereas in the case of a temporal offset, each pixel in the output frame is  $1/2:1/2$  interpolated between pixels or pixel patches in preceding and succeeding progressive format frames, with spatial offset between the  
15 pixels or patches in the source frames and the pixel in the output frame being dependent upon a motion vector which is developed for that pixel.

In the case where there is no temporal offset and a pixel at location  $(x,y)$  in the output frame has a motion vector  $(m,n)$ , this  
20 pixel is derived from the pixel at location  $(x,y)$  or a patch centred on  $(x,y)$  in an input frame 1 to an interpolator, and the motion vector and the content of an input frame 2 are not employed. Also, in the case where there is a half field period temporal offset and an even motion vector  $(m,n)$  for an output pixel at location  $(x,y)$ , the value of this  
25 pixel is derived by equal interpolation between the pixel at (or patch centred on) location  $(x,y) - 1/2(m,n)$  in input frame 1 and the pixel/patch at location  $(x,y) + 1/2(m,n)$  in input frame 2. However, in the case where either of the components of the motion vector  $(m,n)$  is an odd integer, the required pixels or patches in the input frames 1  
30 and 2 will be at locations which are half-way between actual pixel positions in the input frames.

In order to acquire the required pixel values from the input frames, a two-dimensional patch may be used around the required pixel

location, and there will therefore be an offset of (0,0), (1/2,0), (0,1/2), or (1/2,1/2) between the centre pixel of the patch and the pixel location determined by the interpolator depending upon whether neither, one or the other, or each of the motion vector components is  
5 odd. To determine the value of the required pixel, spatial interpolation coefficients are applied to the pixels in the patch, and the sets of coefficients may be chosen to be slightly different for the four possible offsets, although the coefficients for the cases of offset (1/2,0) and (0,1/2) may be symmetrical about the x=y diagonal of  
10 the patch.

A problem which can arise with such an arrangement is that the magnitude responses for the four different sets of spatial interpolation coefficients can be different and produce modulation of the picture detail as the different responses are cycled.

15 One aspect of the present invention aims to deal with this problem of undesirable picture detail modulation.

In accordance with said one aspect of the present invention, there is provided a method of forming an output image by motion compensated interpolation between a pair of input images, comprising  
20 the steps, for each pixel location in the output image, of:

producing a respective motion vector;

determining respective source locations in the input images in dependence upon the respective motion vector; and

calculating the value of the respective pixel in the output image  
25 from respective values derived for the pixels at the source locations in the input images;

wherein the source locations are displaced from the location in the output image by amounts proportional to the respective motion vector plus a constant amount. In an ideal case, the constant amount  
30 is one quarter of a pixel location spacing in each dimension and this will result in the possible offsets being (+ or - 1/4, + or - 1/4).

Therefore, all of the possible offsets are rotationally symmetrically, and thus picture detail modulation can be reduced or



obviated. This method is of particular advantage when the output image is to be interpolated half-way between the input images, and in this case the constants of proportionality are preferably chosen to be equal in magnitude and opposite in sign for the respective two source  
5 locations.

In the arrangement described in GB 8909656.4, with conversion from 60 field/s 2:1 interlace HDVS to 24 Hz 1:1 film format, every other output frame from the interpolator 48 is produced from one of the progressive frames input to the interpolator, and the alternate output  
10 frames are produced by motion compensation between two progressive frames input to the interpolator. This can result in (a) perspective changes not being satisfactorily merged, (b) alias effects when the progressive scan conversion fails due to noise, and (c) noise level modulation when the input image is noisy. As regards point (b), when  
15 progressive scan conversion fails due to noise, the progressive scan frames are produced by intrafield interpolation. If such a frame is directly output by the interpolator, stationary images would appear heavily aliased.

A second aspect of the present invention aims to deal with this  
20 problem, and in accordance with said second aspect of the invention, there is provided a method of forming a series of output images from a series of input images in which some of the output images are temporally aligned with respective input images and some of the output images are temporally offset from the input images, comprising the  
25 steps of:

producing a respective motion vector for each pixel location in each output image;

in the case of a temporally offset output image, determining, for each pixel location therein, respective source locations in the  
30 temporally preceding and temporally succeeding input images which are displaced from the pixel location in the output image by amounts dependent upon the respective motion vector and the temporal offset;

in the case of a temporally aligned output image, determining,

for each pixel location therein, a respective source location in the temporally aligned input image, and a respective source location in the preceding or succeeding input image which is displaced from the pixel location in the output image by an amount dependent upon the respective motion vector; and

calculating for each pixel location in each output image the value of the pixel from respective values derived for the pixels at the two source locations in the respective input images. By producing each pixel in the output frame from two input frames, whether or not there is a temporal offset between the output frame and the input frames, alias is removed, because the frame 2 alias will always be in antiphase to the frame 1 alias as long as the interfield motion is an exact multiple of lines. As synthesised lines are mixed with non-synthesised lines in this scheme, an improved vertical response is also produced. A further advantage is that if there is noise in the input image, the noise will not be modulated, unlike the case where every other output frame is derived from only one of the input frames.

It will be appreciated that the methods of both aspects of the invention may be combined in the same method.

In the case of either method, or the combined method, the values derived for the pixels at the source locations are preferably derived by spatial interpolation between a set of pixel values around the source location.

According to a third aspect of the present invention, there is provided an apparatus adapted to perform the method of the first and/or second aspect of the invention.

The invention will now be described by way of example with reference to the accompanying drawings (and in particular Figures 61 to 65), throughout which like parts are referred to by like references, and in which:

Figure 1 is a block diagram of a previously proposed apparatus for video signal to photographic film conversion;

Figure 2 is a block diagram of part of an embodiment of apparatus for video signal to photographic film conversion according to the present invention;

Figure 3 is a block diagram of another part of the embodiment;

5        Figure 4 is a more detailed block diagram of part of the embodiment;

Figure 5 shows diagrammatically progressive scan conversion;

Figures 6 to 9 show diagrammatically sequences of lines in sequences of fields for explaining progressive scan conversion;

10       Figure 10 is a block diagram showing the steps in motion adaptive progressive scan conversion;

Figure 11 shows diagrammatically progressive scanning, in particular the required estimate and difference value between successive fields;

15       Figures 12 and 13 are diagrams used in explaining the technique of Figure 11 in more detail, Figure 12 showing a progressive scan normalizing function and Figure 13 showing a progressive scan non-linear function;

Figure 14 shows diagrammatically the creation of pixels in missing lines in progressive scan conversion;

20       Figure 15 and 16 show diagrammatically search blocks and search areas, and the relationships therebetween;

Figure 17 shows a correlation surface;

25       Figures 18 and 19 show diagrammatically how a search block is grown;

Figure 20 shows the areas of a frame in which search block matching is not possible;

Figure 21 shows diagrammatically a moving object straddling three search blocks;

30       Figures 22 to 24 show three resulting correlation surfaces, respectively;

Figures 25 and 26 show further examples of correlation surfaces, used in describing a threshold test;

Figures 27 and 28 show still further examples of correlation surfaces, used in describing a rings test;

Figure 29 shows diagrammatically how the direction in which a search block is to grow is determined;

5       Figure 30 shows diagrammatically how a correlation surface is weighted;

Figure 31 shows the relationship between sample blocks and search blocks, and a frame of video;

Figure 32 shows motion vector regions in a frame of video;

10       Figures 33 to 35 show diagrams used in explaining motion vector reduction in respective regions of a frame of video;

Figures 36 and 37 show diagrammatically a first stage in motion vector selection;

15       Figures 38 and 39 show diagrammatically how a threshold is established during the motion vector selection;

Figure 40 shows diagrammatically a second stage in motion vector selection;

Figures 41 to 47 show arrays of pixels with associated motion vectors, used in explaining motion vector post-processing; and

20       Figure 48 shows diagrammatically the operation of an interpolator.

Figure 49 is a diagram illustrating the correlation between frames of a 24 Hz 1:1 format signal and a 60 field/s 3232 pulldown format signal;

25       Figure 50 is a diagram illustrating signal conversion from 60 field/s 2:1 interlace format to 60 field/s 3232 pulldown format;

Figure 51 shows a modification to part of Figure 50;

30       Figure 52 is a diagram illustrating a basic correlation between frames of a 30 Hz 1:1 format signal and fields of a 60 field/s 2:1 interlace format signal with interpolation of alternate output fields only;

Figure 53 shows a modification to Figure 52 with interpolation of all output fields;

Figure 54 is a diagram illustrating signal conversion from 30 Hz 1:1 format to 60 field/s 2:1 interlace format;

Figure 55 is a diagram illustrating a correlation between frames of a 24 Hz 1:1 format signal and fields of a 60 field/s 2:1 interlace format signal;

Figure 56 is a diagram illustrating signal conversion from 24 Hz 1:1 format to 60 field/s 2:1 interlace format;

Figure 57 is a diagram illustrating the source fields of a 60 field/s 2:1 interlace format signal used to produce each frame of a 30 Hz 1:1 format signal with motion adaptive progressive scan conversion only;

Figure 58 is a diagram illustrating the signal conversion of Figure 57;

Figure 59 shows a modification to Figure 57 with motion compensation interpolation also;

Figure 60 is a diagram illustrating the signal conversion of Figure 59;

Figure 61A to 61C show three examples of the relation between a pixel in an output frame and the respective source pixels in two input frames of the motion compensation interpolator of Figure 48;

Figures 62A to 62D shows the different four possible offsets between the location of a required pixel in a input frame shown in Figures 61A to 61D and the actual pixel positions in the input frame;

Figures 63A to 63C are similar to Figures 61A to 61C, respectively, but showing the case where a global offset of (-1/4 pixel, -1/4 pixel) is applied;

Figures 64A to 64D are similar to Figure 62A to 62D, respectively, but showing the offset of Figures 63A to 63C;

Figure 65 is similar to a combination of Figures 63A and 64A, but showing how a pixel in an output frame can be derived from pixels in both input frames, even when there is no temporal offset;

Figure 66 illustrates an overall system primarily for transfer from 24 Hz 1:1 film to 24 Hz 1:1 film and allowing post production

integration with 60 field/s 2:1 interlaced format material;

Figure 67 illustrates an overall system primarily for transfer from 24 Hz 1:1 film to 60 field/s 2:1 interlace HDVS and allowing post production integration with 60 field/s 2:1 interlaced format material;

5      Figure 68 illustrates an overall system primarily for transfer from 30 Hz 1:1 film to 30 Hz 1:1 film and allowing post production integration with 60 field/s 2:1 interlaced format material and 30 Hz 1:1 format material;

10      Figure 69 illustrates an overall system primarily for transfer from 30 Hz 1:1 film and 60 Hz 1:1 film to 60 field/s 2:1 interlace HDVS and allowing post production integration with 60 field/s 2:1 interlaced format material.

15      Figure 70 illustrates an overall system primarily for transfer from 30 Hz 1:1 film to 24 Hz 1:1 or 30 Hz 1:1 film and allowing post production integration with 30 Hz 1:1 video or 60 Hz 2:1 video;

Figure 71 shows a modification to Figure 2 to increase the conversion rate;

20      Figure 72 shows a modification to Figure 2 to enable conversion to take place in two phases requiring less starting and stopping of the video tape recorders;

Figure 73 shows a modification to Figure 2 which obviates the need for a slow-motion source video tape recorder;

Figure 74 shows a modification to Figure 73 which also increases the conversion rate;

25      Figures 75 to 78 illustrates how an image area may be sub-divided into intermediate areas for determining and applying intermediate motion vectors; and

Figure 79 illustrates schematically how local, global and various intermediate motion vectors are made available for selection.

The embodiment of apparatus for video signal to photographic film conversion to be described is particularly intended for use in the conversion of a high definition video signal (HDVS) having 1125 lines per frame, 60 fields per second, to 24 frames per second 35mm film. However, it will be understood that the invention is not limited in this respect, and that it can readily be adapted to effect conversion from other input video signals.

The apparatus can conveniently be considered in two parts; the first part, shown in Figure 2, effects the conversion of the input HDVS to a progressive scan digital video signal corresponding to 24 frames per second which is recorded on a VTR; and the second part, shown in Figure 3, reproduces the recorded video signal and transfers it to photographic film.

The part of the apparatus shown in Figure 2 comprises a high definition digital VTR 11, a television standards converter 12, a frame recorder 13 which can record up to say one second of video signal, a second high definition digital VTR 14, a system controller 15 having associated with it a tracker ball control 16, a keyboard 17 and a graphics display 18, and television monitors 19 and 20, interconnected as shown, and operating as will be described below.

The second part of the apparatus, shown in Figure 3, comprises a high definition digital VTR 31, a digital interface (I/F) unit 32, a gamma corrector 33, a digital-to-analogue converter 34, an electron beam recorder 35, a television monitor 36 and a switch 37, interconnected as shown, and operating as will be described below.

Referring again to Figure 2, the video signal connections D are digital connections, that is carrying Y, U/V signals, and the video signal connections A are analogue connections carrying R, G, B signals. The input video signal which is to be transferred to film, and which may have been derived from a high definition video camera, is recorded on a magnetic tape reproduced by the digital VTR 11. The digital VTR 11 is capable of reproducing the recorded video signal at 1/8 speed, as this is a convenient speed of operation for the subsequence circuitry, and in particular the standards converter 12. The elements 11 to 14, 19 and 20 are under control of the system controller 15, the system controller 15 being in turn controllable by inputs from the tracker ball control 16 and the keyboard 17, and having associated with it the

graphics display 18 on which is displayed information relating to the progress of the conversion.

5 A portion of the input HDVS is reproduced from the digital VTR 11 and supplied to the standards converter 12. This operates, as described in detail below, to derive from the input video signal, which is a 60 fields per second interlace scanned video signal, firstly, a motion adapted progressive scan digital video signal at 60 frames per second, and then from this the required motion compensated progressive scan digital video signal corresponding to 24 frames per second, but  
10 not necessarily at that rate. This video signal is recorded by the digital VTR 14, and if the digital VTR 14 is capable of recording in slow motion, that is at the reproduction rate of the digital VTR 11, then in theory the frame recorder 13 is not required. In practice, however, the frame recorder 13 may in any case be a useful addition to  
15 the apparatus, as it more readily permits intermittent operation to be effected. Such intermittent operation is generally required for video signal to film conversion, because of the need to check at frequent intervals that the conversion is proceeding satisfactorily. Thus depending on the content of the video signal to be converted, adjustment of the parameters, in particular those of the standards  
20 converter 12, need to be made, and the results evaluated before proceeding. The monitors 19 and 20 are provided as further means for checking the video signal at respective points in the apparatus.

In the second part of the apparatus, shown in Figure 3, the  
25 motion compensated progressive scan digital video signal recorded by the digital VTR 14 (Figure 2) is reproduced by the digital VTR 31 and passed by way of the digital I/F unit 32 to the gamma corrector 33, the purpose of which is to match the gamma characteristics of the video signal to the gamma characteristics of the film being used. The  
30 separated operation permitted by the recording of the motion compensated progressive scan digital video signal by the digital VTR 14 (Figure 2), for subsequent reproduction by the digital VTR 31, enables the gamma correction to be set accurately by the gamma corrector 33, because intermittent and repeated operation is possible so that various  
35 different mappings of the generally non-linear gamma characteristics of the video signal from the digital VTR 31 to the generally linear gamma characteristics of the film can be tested. This gamma setting may, for



example, involve the use of a step wedge. The gamma corrected digital video signal is then converted to an analogue signal by the digital-to-analogue converter 34 and supplied to the electron beam recorder 35 to be recorded on photographic film. This recording may, for example, be  
5 in the form of three monochrome frames for each frame of the video signal, the three frames corresponding respectively to red, green and blue. The further television monitor 36 can be selectively connected by way of the switch 37 to the output of the digital VTR 31 or to the output of the digital-to-analogue converter 34, or alternatively of  
10 course two separate television monitors can be provided.

The characteristics of the apparatus are such that it produces sharp, clear pictures with good motion portrayal on the film, and in particular it produces pictures without motion blur and without introducing any additional judder components. Moreover, the separated  
15 operation permitted by the recording of the motion compensated progressive scan digital video signal on the digital VTR 14, in turn permits easy and frequent checking of the parameters of the apparatus, to ensure the quality of the pictures obtained on the film. Iterative operation is perfectly possible, so that the results can rapidly be  
20 evaluated and conversion repeated with any flaws corrected by adjustment of the parameters. To obtain higher speed operation, it is of course possible for the first part of the apparatus, that is the part shown in Figure 2 to be replicated a number of times, to provide additional inputs to the digital VTR 31, so permitting a more intensive  
25 use of the part of the apparatus shown in Figure 3, and hence a higher overall conversion speed.

Figure 4 is a block diagram of the standards converter 12 which will now be described in more detail. The standards converter 12 comprises an input terminal 41 to which an input video signal is  
30 supplied. The input terminal is connected to a progressive scan converter 42 in which the input video fields are converted into video frames which are supplied to a direct block matcher 43 wherein correlation surfaces are created. These correlation surfaces are analysed by a motion vector estimator 44, which derives and supplies  
35 motion vectors to a motion vector reducer 45, wherein the number of motion vectors for each pixel is reduced, before they are supplied to a motion vector selector 46, which also receives an output from the

progressive scan converter 42. Any irregularity in the selection of the motion vectors by the motion vector selector 46 is removed by a motion vector post processor 47, from which the processed motion vectors are supplied to and control an interpolator 48 which also  
5 receives an input from the progressive scan converter 42. The output of the interpolator 48, which is a standards-converted and motion-compensated video signal is supplied to an output terminal 49. Each part of the standards converter 12 and the operation thereof will be described in more detail below.

10 The progressive scan converter 42 produces output frames at the same rate as the input fields. Thus, referring to Figure 5 which shows a sequence of consecutive lines in a sequence of consecutive fields, the crosses representing lines present in the input fields and the squares representing interpolated lines, each output frame will contain  
15 twice the number of lines as an input field, the lines alternating between lines from the input video signal and lines which have been interpolated by one of the methods to be described below. The interpolated lines can be regarded as an interpolated field of the opposite polarity to the input field, but in the same temporal  
20 position.

Progressive scan conversion is preferably carried out, for two main reasons; firstly, to make the following direct block matching process easier, and secondly in consideration of the final output video format. These two reasons will now be considered in more detail.

25 Direct block matching is used to obtain an accurate estimation of the horizontal and vertical motion between two successive video fields, as described in more detail below. However, due to the interlaced structure of the video signal on which direct block matching is performed, problems can arise.

30 Consider the image represented by Figure 6, which indicates a sequence of successive lines in a sequence of successive fields, the open squares representing white pixels, the black squares representing black pixels, and the hatched squares representing grey pixels. This, therefore, represents a static picture with a high vertical frequency component which in a HDVS would be  $1125/3$  cycles per picture height.  
35 As this image has been sampled by the usual interlace scanning procedure, each field appears to contain a static vertical frequency

luminance component Y of 1125/6 cph, as indicated in Figure 7. However, the frequency components in each field are seen to be in anti-phase. Attempts to perform direct block matching between these two fields will lead to a number of different values for the vertical motion component, all of which are incorrect. This is indicated in Figure 8, in which the abbreviation LPF means lines per field. From Figure 8 it is clear that direct block matching will not give the correct answer for the vertical motion component, which component should in fact be zero. This is because the direct block matching is in fact tracking the alias component of the video signal rather than the actual motion.

Consider now Figure 9, which depicts the same static image as Figure 6, except that now each input field has been progressive scan converted to form a frame, the triangles representing interpolated pixels. It can be seen that each frame now contains the same static vertical frequency component as the original input fields, that is 1125/3 cph. Thus, direct block matching between two successive frames can now give the correct value for the vertical motion, that is, zero, and the tracking of the vertical alias has been avoided. Moreover, there is the point that direct block matching on progressive scan converted frames will result in a more accurate vertical motion estimate, because the direct block matching is being performed on frames which have twice the number of lines.

Concerning consideration of the final output video format, in the case of the present embodiment, the converted video is supplied via tape to an electron beam recorder, and needs to consist of frames corresponding to the motion picture film rate of 24 frames per second. For this reason, therefore, the production of progressive scan converted frames is necessary, and moreover the progressive scan converted frames can also be used as a fall-back in the case where motion compensated standards conversion is deemed to be producing unacceptable results, for example, where the motion is too diverse to be analysed satisfactorily. In that case the use of the nearest progressive scan converted frame as the required output frame can produce reasonably acceptable results.

Progressive scan conversion can be carried out in a number of ways, such as by previous field replacement, median filtering in which

three spatially consecutive lines are examined (temporally these three lines will come from two consecutive fields), or a motion compensated technique which utilizes multi-gradient motion detection followed by multi-direction linear interpolation. However, in the present  
 5 embodiment the preferred method is motion adaptive progressive scan conversion, the steps of which are indicated in the block diagram of Figure 10. The concept is to use inter-field interpolation in wholly static picture areas to retain as much vertical information as possible, and to use intra-field interpolation when significant motion  
 10 is present. This also aids smooth portrayal of motion. In scenes where the motion is somewhere between these two extremes, an estimate of the local motion present in the picture is made, and this is then used to mix together different proportions of inter- and intra-field interpolation.

15 In more detail, the modulus of the frame difference between previous and next fields is first generated, this being indicated in Figure 11. To generate the required estimates, the modulus inter-frame difference array from the previous and the next fields is generated at each point:

$$\begin{aligned} \Delta_U(\text{pixel}, \text{current line}, \text{current field}) = \\ |Y(\text{pixel}, \text{current line}, \text{next field}) - \\ Y(\text{pixel}, \text{current line}, \text{previous field})| \end{aligned}$$

where:

$\Delta_U$  is the unnormalized modulus difference array, and  
 25  $Y$  is the luminance array corresponding to the 3D picture.

The modulus of difference is then normalized to adjust for the significance of changes in lower luminance areas:

$$\begin{aligned} \Delta_N(\text{pixel}, \text{current line}, \text{current field}) = \\ F(\overline{Y}(\text{pixel}, \text{current line})) * \\ \Delta_U(\text{pixel}, \text{current line}, \text{current field}) \end{aligned}$$

where:

$\Delta_N$  is the normalized modulus difference array

$\overline{Y}$  is the inter-frame average luminance value

$\overline{Y}(\text{pixel}, \text{current line}) =$

35  $(Y(\text{pixel}, \text{current line}, \text{previous field}) +$

$Y(\text{pixel}, \text{current line}, \text{next field}) )/2$ , and

$F(\overline{Y})$  (the normalizing function) is derived as indicated in Figure

12.

The difference array  $\Delta$  is then vertically filtered together with the previous field difference by a three-tap filter (examples of coefficients are a quarter, a half, a quarter or zero, unity, zero) to  
 5 reduce vertical alias problems, and in particular to minimize the problems encountered with temporal alias. Thus:

$$\begin{aligned} \Delta_P(\text{pixel}, \text{current line}, \text{current field}) = & \\ & \Delta_N(\text{pixel}, \text{current line}-1, \text{previous field}) * C_1 + \\ & \Delta_N(\text{pixel}, \text{current line}, \text{current field}) * C_2 + \\ 10 & \Delta_N(\text{pixel}, \text{current line}+1, \text{previous field}) * C_1 \end{aligned}$$

where:

$\Delta_P$  is the filtered normalized difference array, and

$C_1$  and  $C_2$  are filter coefficients, and  $2C_1 + C_2 = 1$  so that unity dc gain is maintained.

15 A vertical and horizontal intra-field filter of up to five taps by fifteen taps is then used to smooth the difference values within the current field. In practice, a filter of three taps by three taps is satisfactory. Finally, in order to produce the actual motion estimation, a non-linear mapping function is applied using a function  
 20 to provide the motion estimate (ME):

$$\begin{aligned} \text{ME}(\text{pixel}, \text{current line}) = & \\ & \gamma(\text{spatially filtered } \Delta_P(\text{pixel}, \text{current line})) \end{aligned}$$

The non-linear function  $\gamma$  is derived as shown in Figure 13, the static picture ME is zero, for full motion ME is one, and for intermediate  
 25 motions a controlled transition occurs.

To produce an interpolated pixel, the pixels in the missing line are created by taking proportions of the surrounding lines as indicated in Figure 14. The motion estimate ME is then applied to the intra-frame interpolated value (generated from a two, four, six or preferably  
 30 eight tap filter), and  $1-\text{ME}$  is applied to the inter-field average (or alternatively to a more complex interpolated value), and these are summed to derive the progressive scan pixel estimate:

$$\begin{aligned} Y_{\text{out}}(\text{pixel}, \text{current line}) = & \\ & \text{ME}(\text{pixel}, \text{current line}) * \\ 35 & \left\{ \sum_{n=0 \text{ to } 3} (Y_{i_n}(\text{pixel}, \text{current line}-1-2n, \text{current field}) + \right. \\ & \left. Y_{i_n}(\text{pixel}, \text{current line}+1+2n, \text{current field})) * C_n \right\} + \\ & (1-\text{ME})(\text{pixel}, \text{current line}) * \end{aligned}$$

$$( Y_{i0} (\text{pixel, current line, previous field}) + \\ Y_{i0} (\text{pixel, current line, next field}) )/2$$

where:

$C_0, C_1, C_2$  and  $C_3$  are the intra-frame filter coefficients, and  
5  $2(C_0+C_1+C_2+C_3) = 1$  so that unity dc gain is maintained.

This method of progressive scan conversion is found to produce high quality frames from input fields, in particular because a moving object can be isolated and interpolated in a different manner to a stationary background.

10 Referring back to Figure 4, the frames of video derived by the progressive scan converter 42 are used to derive motion vectors. The estimation of motion vectors consists of two steps. Firstly, correlation surfaces are generated by correlating search blocks from consecutive frames. Then, having obtained these correlation surfaces,  
15 they have to be examined to determine the position or positions at which correlation is best. Several different methods of obtaining a correlation surface exist, the two main methods being phase correlation and direct block matching. There are, however, a number of problems associated with the use of phase correlation, these being very briefly  
20 problems relating to the transform mechanism, the windowing function, the block size and the variable quality of the contour of the surface produced. In the present embodiment, therefore, direct block matching is preferred.

The direct block matcher 43 operates as follows. Two blocks,  
25 respectively comprising a rectangular array of pixels from consecutive frames of the progressive scan converted video signal are correlated to produce a correlation surface from which a motion vector is derived.

Referring to Figure 15, firstly a small block called a search block of size 32 pixels by 23 lines is taken from a frame as shown in  
30 Figure 15. Then a larger block called a search area of size 128 pixels by 69 lines is taken from the next frame. The search block (SB) is then placed in each possible position in the search area (SA) as shown in Figure 16, and for each location the sum of the absolute difference of pixel luminance levels between the two blocks is calculated. This  
35 value is then used as the height of the correlation surface at the point at which it was derived. It can then be used in conjunction with other similarly derived values for each possible location of the search

block in the search area to obtain a correlation surface, an example of which is shown in Figure 17. For clarity the surface is shown inverted, and as it is in fact the minimum that is required, the required point in Figure 17 is the main peak.

5       The size of the search block is selected by examining the minimum size of an object that may require motion compensation. For PAL 625 lines per frame, 50 fields per second signals a search block of 16 pixels by 8 lines has been found suitable for tracking a small object without allowing any surrounding information not within the object, but  
10 still within the search block, to affect the tracking of the object. This approach has therefore been adopted in the present embodiment, but modified to take account of the different numbers of active pixels per line, active lines per frame, and aspect ratio of a HDVS as compared with PAL 625/50. The comparative figures, the HDVS being put first,  
15 are as follows; 1920 (720) active pixels per line, 1035 (575) active lines per frame, 3:5.33 (3:4) aspect ratio.

It should be added that there is an argument for using a larger search block, since this means that a large object can be tracked. On the other hand, there exists an argument for using a smaller search  
20 block, to prevent a small object being over-shadowed by the effect of a large object or background area. Also, however, there is the advantage that with small search blocks there is no requirement for the derivation of more than one motion vector from each of them. Because having a single motion vector is so much easier than having more than  
25 one, the present embodiment starts with a small search block as described above, and then causes the search block to grow into a bigger search block if no satisfactory result has been obtained. This then encompasses the advantages of both a small and a large search block. The criteria for a satisfactory result is set by the motion vector  
30 estimator 44 (Figure 4) referred to in more detail below and which determines the motion vector from a given correlation surface.

This technique of causing the search block to grow is not only advantageous for tracking large objects. It can also help to track the movement of an object having the shape of a regular pattern of a  
35 periodic nature. Thus, consider Figure 18 where a search block A will match up with the search area B at locations V1, V2 and V3, with each of them giving a seemingly correct measure of motion. In this case,

however, the motion vector estimation, that is the process that actually analyses the correlation surface, will show that good correlation occurs in three locations which are collinear. The search block will therefore be caused to grow horizontally until it is three times its original width, this being the direction in which multiple correlation occurred in this case. The search area will also be correspondingly horizontally enlarged. As shown in Figure 19, with the enlarged search block 3A, there is only a single correlation point, which correctly relates to the motion of the object.

In this particular case the search block and the search area both have to grow horizontally, because the direction of multiple correlation is horizontal. It is equally possible, however, for the search block and the search area to grow vertically, or indeed in both directions, if the correlation surface suggests it.

It should be noted that block matching cannot be applied to all the search blocks in the frame, because in the border area there is not enough room from which a search area can be drawn. Thus, block matching cannot be effected in the border area of the frame shown hatched in Figure 20. This problem is dealt with by the motion vector reducer 45 (Figure 4) described in more detail below, which attempts to supply search blocks in this hatched area with appropriate motion vectors.

From the correlation surface (Figure 17) generated for each search block in a frame the motion vector estimator 44 (Figure 4) deduces the likely inter-frame motion between the search block and its corresponding search area. It should again be mentioned that for clarity all diagrams of correlation surfaces are shown inverted, that is, such that a minimum is shown as a peak.

The motion vector estimator 44 (Figure 4) uses motion vector estimation algorithms to detect the minimum point on each correlation surface. This represents the point of maximum correlation between the search block and the search area, and hence indicates the probable motion between them. The displacement of this minimum on the correlation surface with respect to the origin, in this case the centre of the surface, is a direct measurement, in terms of pixels per frame, of the motion. For the simplest case, where the correlation surface contains a single, distinct minimum, the detection of the minimum point



on the correlation surface is sufficient to determine accurately the motion between the search block and the search area. As previously mentioned, the use of small search blocks improves the detection of motion and the accuracy of motion estimation, but unfortunately small  
5 single search blocks are unable to detect motion in a number of circumstances which will now be described.

Figure 21 shows an object with motion vectors (5, 0) straddling three search blocks 1A, 2A and 3A in a frame (t). When the search blocks 1A and 3A are correlated with respective search areas (1B and  
10 3B) in the next frame (t+1) a correlation surface shown in Figure 22 results showing a minimum at (5, 0). (This assumes a noiseless video source.) However, when the search block 2A is correlated with its respective search area 2B, the correlation surface shown in Figure 23 is produced, in which the search block 2A correlates with the search  
15 area 2B at every point in the y-axis direction. There is therefore no single minimum in the correlation surface, and hence the motion between the search block 2A and the search area 2B cannot be determined.

However, now consider the situation if the search block 2A is grown such that it encompasses all three of the original search blocks  
20 1A, 2A and 3A. When the grown search block 2A is correlated with a search area covering the original search areas 1B, 2B and 3B, the resulting correlation surface is as shown in Figure 24. This shows a single minimum at (5, 0) indicating the correct motion of the original search block 2A. This example illustrates the need for some unique  
25 feature in the source video, in order accurately to detect motion. Thus, the search blocks 1A and 3A both had unique vertical and horizontal features, that is the edges of the object, and hence motion could be determined. In contrast, the search block 2A had a unique vertical feature, but no unique horizontal feature, and hence  
30 horizontal motion could not be determined. However, by growing the search block until it encompasses a unique feature both horizontally and vertically, the complete motion for that search block can be determined. Moreover, it can be shown that growing the search block is beneficial when noise in the source video is considered.

35 A further example will now be considered with reference to Figure 25. This shows a correlation surface for a search block where the motion vector is (5, 3). However, due to the numerous other

correlations which have taken place between the search block and the search area, the true motion is difficult to detect. An example of source video which might produce such a correlation surface would be a low contrast tree moving with the wind. It is now assumed that the search block and the search area are grown. The growing can take place in the horizontal direction, as in the previous example, or in the vertical direction, or in both directions. Assuming that the neighbouring search blocks have the same motion, the mean effect on the resulting correlation surface will be to increase the magnitude of the minima at (5, 3) by a greater proportion than the magnitude of the other correlation peaks. This is shown in Figure 26, which indicates that it is then easier to detect the correct motion vector.

The way in which search blocks are grown will now be further considered with reference to Figure 21. Here it was required to grow the area of the search block 2A to encompass the areas of the search blocks 1A and 3A, and to produce the resulting correlation surface. In fact, the resulting correlation surfaces are produced directly by adding together the elements of the three correlation surfaces corresponding to the search blocks 1A, 2A and 3A. In effect, if each correlation surface is considered as a matrix of point magnitudes, then the correlation surface of the enlarged search block 2A is the matrix addition of the correlation surface of the original search blocks 1A, 2A and 3A.

The area of the search block 2A could also be grown vertically by adding correlation surfaces of the search blocks above and below, whilst if the search block 2A is to be grown both horizontally and vertically, then the four neighbouring diagonal correlation surfaces have to be added as well. From this it will be seen that the actual process of growing a search block to encompass neighbouring search blocks is relatively easy, the more difficult process being to decide when growing should take place, and which neighbouring search blocks should be encompassed. Basically, the answer is that the area of the search blocks should be grown until a good minimum or good motion vector is detected. It is therefore necessary to specify when a motion vector can be taken to be a good motion vector, and this can in fact be deduced from the examples given above.

In the example described with reference to Figures 21 to 24, it

was necessary to grow the search block horizontally in order to encompass a unique horizontal feature of the object, and hence obtain a single minimum. This situation was characterized by a row of identical minima on the correlation surface of Figure 23, and a single minimum on the correlation surface of Figure 24. From this the first criteria for a good minimum can be obtained; a good minimum is the point of smallest magnitude on the correlation surface for which the difference between it and the magnitude of the next smallest point exceeds a given value. This given value is known as the threshold value, and hence this test is referred to herein as the threshold test.

It should be noted that the next smallest point is prevented from originating from within the bounds of a further test, described below, and referred to herein as the rings test. In the case of a rings test employing three rings, the next smallest point is prevented from originating from a point within three pixels of the point in question. In the example of Figures 21 to 24, the correlation surface of Figure 23 would have failed the threshold test; the search area 2A is therefore grown and, given a suitable threshold value, the correlation surface of Figure 24 will pass the threshold test.

The threshold test can also be used to cause growing in the example described above with reference to Figures 25 and 26. Prior to growing the search block, the correct minimum is undetectable, due to the closely similar magnitudes of the surrounding points. Given a suitable threshold value, however, the correlation surface will fail the threshold test, and the search block will be grown. As a result, it will then be possible to detect the minimum among the other spurious points.

It will be seen that the use of a threshold is a subjective test, but the correct threshold for the correlation surface under test can be selected by normalizing the threshold as a fraction of the range of magnitudes within the correlation surface. This also lessens the effect of, for example the contrast of the video source.

The rings test, referred to briefly above, and which is far less subjective, will now be further described. The basis of the rings test is to assume that a good minimum (or maximum) will have points of increasing (or decreasing) magnitudes surrounding it. Figure 27 illustrates this assumption, showing a minimum at (0, 0) where the

surrounding three rings of points have decreasing mean magnitude. This is as opposed to the correlation surface shown in Figure 28, where the rings, and in particular the second inner-most ring, are not of decreasing mean magnitude.

5        In this case the criteria for a good minimum as defined by the rings test, is that the average slope is monotonic. Therefore for a pre-defined number of rings of points surrounding the minimum in question, the mean magnitude of each ring when moving from the inner-most ring outwards, must be greater than that of the previous ring.  
10      Returning again to the example described with reference to Figures 21 to 24, it will be seen from Figures 23 and 24 that the correlation surface of Figure 23 would have failed the rings test, but that the correlation surface of Figure 24 would have passed the rings test. Since the rings test compares mean, and not absolute, magnitudes, it is  
15      far less subjective than the threshold test, and indeed the only variable in the rings test is the number of rings considered.

      Having described the mechanism for growing a search block, it is now necessary to consider how by examining the shape of the correlation surface it is possible to determine the most effective direction in  
20      which the search block should grow.

      Referring again to Figure 23, this correlation surface resulted where there was a unique vertical feature, but no unique horizontal feature. This is mirrored in the correlation surface by the minimum running horizontally across the correlation surface, due to the  
25      multiple correlations in this direction. From this it can be deduced that the search block should be grown horizontally. Conversely, should a line of multiple correlations run vertically, this would indicate the need to grow the search block vertically, whilst a circular collection of multiple correlations would indicate a need to grow the search block  
30      both horizontally and vertically.

      Using this criteria, a quantitative measure of the shape of the correlation surface is required in order to determine in which direction the search block should be grown. This measure is determined as follows. Firstly, a threshold is determined. Any point on the  
35      correlation surface below the threshold is then considered. This threshold, like that used in the threshold test, is normalized as a fraction of the range of magnitudes within the correlation surface.

Using this threshold, the points on the correlation surface are examined in turn in four specific sequences. In each, the point at which the correlation surface value falls below the threshold is noted. These four sequences are illustrated diagrammatically in Figure 29 in which the numbers 1, 2, 3 and 4 at the top, bottom, left and right refer to the four sequences, and the hatched area indicates points which fall below the threshold:

Sequence 1

Search from the top of the correlation surface down for a point A which falls below the threshold.

Sequence 2

Search from the bottom of the correlation surface up for a point C which falls below the threshold.

Sequence 3

Search from the left of the correlation surface to the right for a point D which falls below the threshold.

Sequence 4

Search from the right of the correlation surface to the left for a point B which falls below the threshold.

The locations of the four resulting points A, B, C and D are used to calculate the two dimensions X and Y indicated in Figure 29, these dimensions X and Y indicating the size of the hatched area containing the points falling below the threshold value. Hence from the dimensions X and Y, it can be deduced whether the shape is longer in the x rather than the y direction, or vice versa, or whether the shape is approximately circular. A marginal difference of say ten percent is allowed in deducing the shape, that is, the dimension X must be a minimum of ten percent greater than the dimension Y for the shape to be considered to be longer in the x direction. Similarly for the y direction. If the dimensions X and Y are within ten percent of each other, then the shape is considered to be circular, and the search block is grown in both directions. In the example of Figure 29 the dimension X is greater than the dimension Y, and hence the search block is grown in the x or horizontal direction.

The growing of the search block continues until one or more growth limitations is reached. These limitations are: that the minimum in the correlation surface passes both the threshold test and

the rings test; that the edge of the video frame is reached; or that the search block has already been grown a predetermined number of times horizontally and vertically. This last limitation is hardware dependent. That is to say, it is limited by the amount of processing  
5 that can be done in the available time. In one specific embodiment of apparatus according to the present invention, this limit was set at twice horizontally and once vertically.

If the minimum in the correlation surface passes both the threshold test and the rings test, then it is assumed that a good  
10 motion vector has been determined, and can be passed to the motion vector reducer 45 (Figure 4). However, if the edge of the frame is reached or the search block has already been grown a predetermined number of times both horizontally and vertically, then it is assumed that a good motion vector has not been determined for that particular  
15 search block, and instead of attempting to determine a good motion vector, the best available motion vector is determined by weighting.

The correlation surface is weighted such that the selection of the best available motion vector is weighted towards the stationary, that is the centre, motion vector. This is for two reasons, firstly,  
20 if the search block, even after growing, is part of a large plain area of source video, it will not be possible to detect a good motion vector. However, since the source video is of a plain area, a stationary motion vector will lead to the correct results in the subsequent processing. Secondly, weighting is designed to reduce the  
25 possibility of a seriously wrong motion vector being passed to the motion vector reducer 45 (Figure 4). This is done because it is assumed that when a good motion vector cannot be determined, a small incorrect motion vector is preferable to a large incorrect motion vector.

30 Figure 30 shows an example of how the weighting function can be applied to the correlation surface. In this example, the weight applied to a given point on the correlation surface is directly proportional to the distance of that point from the stationary, centre motion vector. The magnitude of the point on the correlation surface  
35 is multiplied by the weighting factor. For example, the gradient of the weighting function may be such that points plus or minus 32 pixels from the centre, stationary motion vector are multiplied by a factor of

three. In other words, as shown in Figure 30, where the centre, stationary motion vector is indicated by the black circle, the weighting function is an inverted cone which is centred on the centre, stationary motion vector.

5       After the correlation surface has been weighted, it is again passed through the threshold test and the rings test. If a minimum which passes both these tests is determined, then it is assumed that this is a good motion vector, and it is flagged to indicate that it is a good motion vector, but that weighting was used. This flag is  
10       passed, together with the motion vector to the motion vector reducer 45 (Figure 4). If on the other hand, neither a good motion vector nor a best available motion vector can be determined, even after weighting, then a flag is set to indicate that any motion vector passed to the motion vector reducer 45 (Figure 4) for this search block is a bad  
15       motion vector. It is necessary to do this because bad motion vectors must not be used in the motion vector reduction process, but must be substituted as will be described below.

      Thus, in summary, the operation of the motion vector estimator 44 (Figure 4) is to derive from the correlation surface generated by the  
20       direct block matcher 43 (Figure 4), the point of best correlation, that is the minimum. This minimum is then subjected to the threshold test and the rings test, both of which the minimum must pass in order for it to be considered to represent the motion of the search block. It should, incidentally, be noted that the threshold used in the threshold  
25       test and the rings test may be either absolute values or fractional values. If the minimum fails either test, then the search block is grown, a new minimum is determined, and the threshold test and the rings test re-applied. The most effective direction in which to grow the search block is determined from the shape of the correlation  
30       surface.

      Referring initially to Figure 4, the process of motion vector reduction will now be described. Using a HDVS, each search block is assumed to be 32 pixels by 23 lines, which can be shown to lead to a possible maximum of 2451 motion vectors. The choice of the search  
35       block size is a compromise between maintaining resolution and avoiding an excessive amount of hardware. If all these motion vectors were passed to the motion vector selector 46, the task of motion vector

selection would not be practicable, due to the amount of processing that would be required. To overcome this problem, the motion vector reducer 45 is provided between the motion vector estimator 44 and the motion vector selector 46. The motion vector reducer 45 takes the motion vectors that have been generated by the motion vector estimator 44 and presents the motion vector selector 46 with only, for example, four motion vectors for each search block in the frame, including those in border regions, rather than all the motion vectors derived for that frame. The effect of this is two-fold. Firstly, this makes it much easier to choose the correct motion vector, so long as it is within the group of four motion vectors passed to the motion vector selector 46. Secondly, however, it also means that if the correct motion vector is not passed as one of the four, then the motion vector selector 46 is not able to select the correct one. It is therefore necessary to try to ensure that the motion vector reducer 45 includes the correct motion vector amongst those passed to the motion vector selector 46. It should also be mentioned that although four motion vectors are passed by the motion vector reducer 45 to the motion vector selector 46, only three of these actually represent motion, the fourth motion vector always being the stationary motion vector which is included to ensure that the motion vector selector 46 is not forced into applying a motion vector representing motion to a stationary pixel. Other numbers of motion vectors can be passed to the motion vector selector 46, for example, in an alternative embodiment four motion vectors representing motion and the stationary motion vector may be passed.

Hereinafter the term 'sample block' refers to a block in a frame of video in which each pixel is offered the same four motion vectors by the motion vector reducer 45. Thus, a sample block is the same as a search block before the search block has been grown. As shown in Figure 31, in a frame of video the initial positions of the sample blocks and the search blocks are the same.

The motion vector reducer 45 (Figure 4) receives the motion vectors and the flags from the motion vector estimator 44 (Figure 4) and determines the quality of the motion vectors by examining the flags. If the motion vector was not derived from an ambiguous surface, that is there is a high degree of confidence in it, then it is termed a good motion vector, but if a certain amount of ambiguity exists, then



the motion vector is termed a bad motion vector. In the motion vector reduction process, all motion vectors classed as bad motion vectors are ignored, because it is important that no incorrect motion vectors are ever passed to the motion vector selector 46 (Figure 4), in case a bad  
5 motion vector is selected thereby. Such selection would generally result in a spurious dot in the final picture, which would be highly visible.

Each of the motion vectors supplied to the motion vector reducer 45 (Figure 4) was obtained from a particular search block, and hence a  
10 particular sample block (Figure 31), the position of these being noted together with the motion vector. Because any motion vectors which have been classed as bad motion vectors are ignored, not all sample blocks will have a motion vector derived from the search block at that position. The motion vectors which have been classed as good motion  
15 vectors, and which relate to a particular search block, and hence a particular sample block, are called local motion vectors, because they have been derived in the area from which the sample block was obtained. In addition to this, another motion vector reduction process counts the frequency at which each good motion vector occurs, with no account  
20 taken of the actual positions of the search blocks that were used to derive them. These motion vectors are then ranked in order of decreasing frequency, and are called common motion vectors. In the worst case only three common motion vectors are available and these are combined with the stationary motion vector to make up the four motion  
25 vectors to be passed to the motion vector selector 46 (Figure 4). However, as there are often more than three common motion vectors, the number has to be reduced to form a reduced set of common motion vectors referred to as global motion vectors.

A simple way of reducing the number of common motion vectors is  
30 to use the three most frequent common motion vectors and disregard the remainder. However, the three most frequent common motion vectors are often those three motion vectors which were initially within plus or minus one pixel motion of each other vertically and/or horizontally. In other words, these common motion vectors were all tracking the same  
35 motion with slight differences between them, and the other common motion vectors, which would have been disregarded, were actually tracking different motions.

In order to select the common motion vectors which represent all or most of the motion in a scene, it is necessary to avoid choosing global motion vectors which represent the same motion. Thus, the strategy actually adopted is first to take the three most frequently occurring common motion vectors and check to see if the least frequent among them is within plus or minus one pixel motion vertically and/or plus or minus one pixel motion horizontally of either of the other two common motion vectors. If it is, then it is rejected, and the next most frequently occurring common motion vector is chosen to replace it. This process is continued for all of the most frequently occurring common motion vectors until there are either three common motion vectors which are not similar to each other, or until there are three or less common motion vectors left. However, if there are more than three common motion vectors left, then the process is repeated this time checking to see if the least frequent among them is within plus or minus two pixel motion vertically and/or plus or minus two pixel motion horizontally of another, and so on at increasing distances if necessary. These three common motion vectors are the required global motion vectors, and it is important to note that they are still ranked in order of frequency.

When considering the motion vector reduction process and the sample blocks of a frame of video, it is necessary to look at three different types of sample blocks. These types are related to their actual position in a frame of video, and are shown in Figure 32 as regions. Region A comprises sample blocks which are totally surrounded by other sample blocks and are not near the picture boundary. Region B contains sample blocks which are partially surrounded by other sample blocks and are not near the picture boundary. Finally, region C contains sample blocks which are near the picture boundary. The motion vector reduction algorithm to be used for each of these regions is different. These algorithms will be described below, but firstly it should be reiterated that there exist good motion vectors for some of the sample blocks in the frame of video, and additionally there are also three global motion vectors which should represent most of the predominant motion in the scene. A selection of these motion vectors is used to pass on three motion vectors together with the stationary motion vector for each sample block.

Figure 33 illustrates diagrammatically motion vector reduction in the region A. This is the most complex region to deal with, because it has the largest number of motion vectors to check. Figure 33 shows a central sample block which is hatched, surrounded by other sample blocks a to h. Firstly, the locally derived motion vector is examined to see if it was classed as a good motion vector. If it was, and it is also not the same as the stationary motion vector, then it is passed on. However, if it fails either of these tests, it is ignored. Then the motion vector associated with the sample block d is checked to see if it was classed as a good motion vector. If it was, and if it is neither the same as any motion vector already selected, nor the same as the stationary motion vector, then it too is passed on. If it fails any of these tests then it too is ignored. This process then continues in a similar manner in the order e, b, g, a, h, c and f. As soon as three motion vectors, not including the stationary motion vector, have been obtained, then the algorithm stops, because that is all that is required for motion vector selection for that sample block. It is, however, possible for all the above checks to be carried out without three good motion vectors having been obtained. If this is the case, then the remaining spaces are filled with the global motion vectors, with priority being given to the more frequent global motion vectors.

Figure 34 illustrates motion vector reduction in the region B. Sample blocks in the region B are the same as those in the region A, except that they are not totally surrounded by other sample blocks. Thus the process applied to these sample blocks is exactly the same as those for the region A, except that it is not possible to search in all the surrounding sample blocks. Thus as seen in Figure 34, it is only possible to check the motion vectors for the sample blocks a to e, and any remaining spaces for motion vectors are filled, as before, with global motion vectors. Likewise, if the hatched sample block in Figure 34 were displaced two positions to the left, then it will be seen that there would only be three adjacent surrounding blocks to be checked before resorting to global motion vectors.

Figure 35 illustrates motion vector reduction in the region C. This is the most severe case, because the sample blocks neither have a locally derived motion vector nor do they have many surrounding sample blocks whose motion vectors could be used. The simplest way of dealing

with this problem is simply to give the sample blocks in the region C the global motion vectors together with the stationary motion vector. However, this is found to produce a block-like effect in the resulting picture, due to the sudden change in the motion vectors presented for the sample blocks in the region C compared with adjoining sample blocks in the region B. Therefore a preferred strategy is to use for the sample blocks in the region C the sample motion vectors as those used for sample blocks in the region B, as this prevents sudden changes. Preferably, each sample block in the region C is assigned the same motion vectors as that sample block in the region B which is physically nearest to it. Thus, in the example of Figure 35, each of the hatched sample blocks in the region C would be assigned the same motion vectors as the sample block a in the region B, and this has been found to give excellent results.

Referring again to Figure 4, the purpose of the motion vector selector 46 is to assign one of the four motion vectors supplied thereto to each individual pixel within the sample block. In this way the motion vectors can be correctly mapped to the outline of objects. The way in which this assignment is effected is particularly intended to avoid the possibility of the background surrounding fine detail from producing a better match than that produced by the correct motion vector. To achieve this the motion vector selection process is split into two main stages. In the first stage, motion vectors are produced for each pixel in the input frames. In other words, there is no attempt to determine the motion vector values for pixels at the output frame positions. The second stage uses the motion vector values produced by the first stage to determine the motion vector value for each pixel in the output frame.

Referring now to Figure 36, each pixel of the input frame 2 is tested for the best luminance value match with the previous and following input frames 1 and 3 of video data, using each of the four motion vectors supplied. The pixel luminance difference is determined as:

$$\sum_{m=0}^4 \sum_{n=0}^4 |P1_{nn} - P2_{nn}| + \sum_{m=0}^4 \sum_{n=0}^4 |P2_{nn} - P3_{nn}|$$

where:

- $P1_{nn}$  is the luminance value of a frame 1 pixel within a 4x4 block of pixels surrounding the pixel whose location is obtained by subtracting the coordinates of the motion vector being tested from the location of the pixel being tested in frame 2
- 5  $P2_{nn}$  is the luminance value of a frame 2 pixel within a 4x4 block of pixels surrounding the pixel being tested
- $P3_{nn}$  is the luminance value of a frame 3 pixel within a 4x4 block of pixels surrounding the pixel whose location is obtained by adding the coordinates of the motion vector being tested to the location
- 10 of the pixel being tested in frame 2

The minimum pixel difference then indicates the best luminance match and therefore the correct motion vector applicable to the pixel being tested. If the correct motion vector is not available, or there are uncovered or covered areas, referred to in more detail below, then

15 a good match may not occur.

The indication of a poor match is achieved when the average pixel difference within the block of pixels being used is above a certain threshold. This threshold is important, because high frequency detail may produce a poor match even when the correct motion vector is tested.

20 The reason for this poor match is the possibility of a half pixel error in the motion vector estimate. To determine what threshold should indicate a poor match, it is necessary to relate the threshold to the frequency content of the picture within the block of data which surrounds the pixel for which the motion vector is required. To

25 achieve this, an auto-threshold value is determined where the threshold value equals half the maximum horizontal or vertical pixel luminance difference about the pixel being tested. To ensure that the threshold value obtained is representative of the whole block of data which is compared, an average value is obtained for the four central pixels of

30 a 4x4 block used.

Referring to Figure 38, which shows a 4x4 block, the required threshold value T is given by:

$$T = (T1 + T2 + T3 + T4)/8$$

where T3, for example, is determined as indicated in Figure 39 as equal

35 to the maximum of the four pixel luminance difference values comprising:

the two vertical differences  $|B2 - B3|$  and  $|B4 - B3|$ , and

the two horizontal differences  $|A3 - B3|$  and  $|C3 - B3|$

In this way a frame of motion vectors is obtained for input frame 2, and in a similar manner a frame of motion vectors is obtained for input frame 3 as indicated in Figure 37.

5        Apart from scene changes, it is the phenomenon of uncovered/covered surfaces that causes a mis-match to occur in the above first stage of motion vector selection. If an object, say a car, drives into a tunnel, then the car has become covered, while when it drives out, the car is uncovered. If the part of the car that was  
10        uncovered in frames 1 and 2 is covered in frames 3 and 4, then the basic vector selection process is not able to determine the correct vector. Moreover, whilst the car going into the tunnel becomes covered, the road and objects behind the car are being uncovered. Likewise the car leaving the tunnel is being uncovered, but the road  
15        and objects behind the car are being covered. In general therefore both covered and uncovered objects will exist at the same time. The end of a scene will also have a discontinuation of motion that is similar to an object becoming covered. In an attempt to determine a motion vector even in such circumstances, the luminance value block  
20        match is reduced to a two frame match, instead of the three frame match of Figures 36 and 37. The frame that the motion vectors are required for (say frame 2) is block-matched individually to the previous and the next frame (frame 1 and frame 3 respectively, in the case of frame 2), using the four motion vectors supplied. The motion vector which  
25        produces the best match is chosen as the motion vector applicable to the pixel being tested. In this case, however, a flag is set to indicate that only a two frame match was used.

Particularly with integrating type television cameras, there will be situations where no match occurs. If an object moves over a  
30        detailed background, then an integrating camera will produce unique portions of picture where the leading and trailing edges of the object are mixed with the detail of the background. In such circumstances, even the two frame match could produce an average pixel difference above the threshold value. In these cases the motion vector value is  
35        set to zero, and an error flag is also set.

The second stage of motion vector selection makes use of the two frames of motion vectors, derived by the first stage. One frame of

motion vectors (input frame 2) is considered to be the reference frame, and the following frame to this (input frame 3) is also used. The output frame position then exists somewhere between these two frames of motion vectors. Referring to Figure 40, for each output pixel position the four possible motion vectors associated with the sample block of input frame 2, are tested. A line drawn through the output pixel position at the angle of the motion vector being tested will point to a position on both the input frame 2 and the input frame 3. In the case of odd value motion vectors, for example, 1, 3 and 5, a point mid-way between two input frame pixels would be indicated in the case where the output frame is precisely half way between the input frames 1 and 2. To allow for this inaccuracy, and also to reduce the sensitivity to individual pixels, a 3x3 block of motion vectors is acquired for each frame, centred on the closest pixel position. In effect a block-match is then performed between each of the two 3x3 blocks of motion vectors and a block containing the motion vector being tested. The motion vector difference used represents the spatial difference of the two motion vector values as given by:

$$\sqrt{((x1-x2)^2 + (y1-y2)^2)}$$

where:

x1 and y1 are the Cartesian coordinates of the motion vector in one of the blocks

x2 and y2 are the Cartesian coordinates of the motion vector being tested

An average vector difference per pixel is produced as a result of the block match.

A motion vector match is first produced as above using only motion vector values which were calculated using three input frames; that is, input frames 1, 2 and 3 for input frame 2 (Figure 36), and input frames 2, 3 and 4 for input frame 3 (Figure 37), and the result is scaled accordingly. Preferably there are at least four usable motion vectors in the block of nine. When both the motion vector block of frame 2 and frame 3 can be used, the motion vector difference values are made up of half the motion vector difference value from frame 2 plus half the motion vector difference value from frame 3. Whichever motion vector produces the minimum motion vector difference value using the above technique is considered to be the motion vector applicable to

the output pixel being tested. If the motion vector difference value produced by the three frame match input motion vector (Figures 36 and 37 is greater than unity, then a covered or uncovered surface has been detected, and the same process is repeated, but this time ignoring the error flags. That is, the motion vector values which were calculated using two input frames are used. Theoretically this is only necessary for uncovered/covered surfaces, although in fact improvements can be obtained to the picture in more general areas.

If after both of the above tests have been performed, the minimum motion vector match is greater than two, the motion vector value is set to zero, and an error flag is set for use by the motion vector post processor 47 (Figure 4).

Following motion vector selection, there will almost certainly be in any real picture situation, some remaining spurious motion vectors associated with certain pixels. Figures 41 to 46 show what are taken to be spurious motion vectors, and in each of these figures the triangles represent pixels having associated therewith the same motion vectors, whilst the stars represent pixels having associated therewith motion vectors different those associated with the surrounding pixels, and the circle indicates the motion vector under test.

Figure 41 shows a point singularity where a single pixel has a motion vector different from those of all the surrounding pixels.

Figure 42 shows a horizontal motion vector impulse, where three horizontally aligned pixels have a motion vector different from those of the surrounding pixels.

Figure 43 shows a vertical motion vector impulse where three vertically aligned pixels have a motion vector different from those of the surrounding pixels.

Figure 44 shows a diagonal motion vector impulse where three diagonally aligned pixels have a motion vector different from those of all the surrounding pixels.

Figure 45 shows a horizontal plus vertical motion vector impulse, where five pixels disposed in an upright cross have a motion vector different from those of all the surrounding pixels.

Figure 46 shows a two-diagonal motion vector impulse where five pixels arranged in a diagonal cross have a motion vector different from those of all the surrounding pixels.



It is assumed that pixel motion vectors which fall into any of the above six categories do not actually belong to a real picture, and are a direct result in of an incorrect motion vector selection. If such motion vectors were used during the interpolation process, then  
5 they would be likely to cause dots on the final output picture, and it is therefore preferable that such motion vectors be identified and eliminated. This is done using an algorithm which will detect and flag all of the above motion vector groupings.

The algorithm uses a two-pass process, with each pass being  
10 identical. The need for two passes will become apparent. Figure 47, to which reference is made, shows an array of pixels, all those marked with a triangle having the same motion vector associated therewith. The block of nine pixels in the centre has motion vectors designated vector 1 to vector 9 associated therewith, which motion vectors may or  
15 may not be the same. Vector 5 is the motion vector under test.

In the first pass, vector 5 is checked to determine whether it is the same as, or within a predetermined tolerance of:

firstly  
vector 1 or vector 3 or vector 7 or vector 9  
20 and secondly  
vector 2 or vector 4 or vector 6 or vector 8

This checks to see if vector 5 is the same as at least one of its horizontal or vertical neighbours, and the same as at least one of its diagonal neighbours. If this is not the case, then a flag to set to  
25 indicate that pixel 5 is bad.

The first pass will flag as bad those motion vectors relating to point singularities, horizontal motion vector impulses, vertical motion vector impulses, diagonal motion vector impulses and two diagonal motion vector impulses (Figures 41 to 44 and 46), but not the motion  
30 vectors corresponding to horizontal plus vertical motion vector impulses (Figure 45) for which pass 2 is required. The second pass checks for exactly the same conditions as in the first pass, but in this case motion vectors which have already been flagged as bad are not included in the calculation. Thus, referring to Figure 45, after the  
35 first pass only the centre motion vector is flagged as bad, but after the second pass all five of the motion vectors disposed in the upright cross are flagged as bad.

Having identified the bad motion vectors, it is then necessary to repair them, this also being effected by the motion vector post processor 47 (Figure 4). Although various methods such as interpolation or majority replacement can be used, it has been found  
5 that in practice simple replacement gives good results. This is effected as follows (and it should be noted that the 'equals' signs mean not only exactly equal to, but also being within a predetermined tolerance of):

If vector 5 is flagged as bad then it is replaced with:

10       vector 4 if (vector 4 equals vector 6)  
          else with vector 2 if (vector 2 equals vector 8)  
          else with vector 1 if (vector 1 equals vector 9)  
          else with vector 3 if (vector 3 equals vector 7)  
          else do nothing

15       Referring again to Figure 4, the finally selected motion vector for each pixel is supplied by the motion vector post processor 47 to the interpolator 48, together with the progressive scan converted frames at 60 frames per second from the progressive scan converter 42. The interpolator 48 is of relatively simple form using only two  
20 progressive scan converted frames, as indicated in Figure 48. Using the temporal position of the output frame relative to successive input frames, frame 1 and frame 2, and the motion vector for the pixel in the output frame, the interpolator 48 determines in known manner which part of the first frame should be combined with which part of the second  
25 frame and with what weighting to produce the correct output pixel value. In other words, the interpolator 48 adaptively interpolates along the direction of movement in dependence on the motion vectors to produce motion compensated progressive scan frames corresponding to 24 frames per second. Although the motion vectors have been derived using  
30 only luminance values of the pixels, the same motion vectors are used for deriving the required output pixel chrominance values. An 8x8 array of pixels are used from each frame to produce the required output. Thus the interpolator 48 is a two-dimensional, vertical/horizontal, interpolator and the coefficients used for the  
35 interpolator 48 may be derived using the Remez exchange algorithm which can be found fully explained in 'Theory and application of digital signal processing', Lawrence R Rabiner, Bernard Gold. Prentice-Hall

Inc., pages 136 to 140 and 227.

Figure 48 shows diagrammatically the interpolation performed by the interpolator 48 (Figure 4) for three different cases. The first case, shown on the left, is where there are no uncovered or covered surfaces, the second case, shown in the centre, is where there is a covered surface, and the third case, shown on the right, is where there is an uncovered surface. In the case of a covered surface, the interpolation uses only frame 1, whilst in the case of an uncovered surface, the interpolation uses only frame 2.

Provision can be made in the interpolator 48 to default to non-motion compensated interpolation, in which case the temporally nearest progressive scan converted frame is used.

In the arrangement described above, particularly with reference to Figure 2, a 60 field/s, 30 frame/s, 2:1 interlaced format signal is converted to 24 frame/s 1:1 progressive format by:

- a) supplying the interlaced signal to the standards converter 12 at one-eighth speed with each input field repeated eight times;
- b) developing, for each input field, a progressive format frame in the progressive scan converter 42;
- c) for each ten repeated input frames, developing an output frame by appropriate interpolation in the interpolator 48 between the two currently developed progressive format frames in dependence upon the supplied motion vector, with the developed output frame being repeated ten times; and
- d) recording one in every ten output frames.

It will therefore be appreciated that the standards converter 12 operates at one-eighth of real-time, and that, for every five input interlaced frames, four output frames will be produced, thus giving the 30 frame/s to 24 frame/s conversion.

In some applications, for example where further processing is to be carried out on the 24 frame/s 1:1 format signal, or where it is desired to record the 24 frame/s format signal on standard HDVS recording equipment and replay it, it is beneficial to use a modified 24 frame/s format employing a "3232 pulldown" sequence. For further description of the 3232 pulldown sequence, reference is directed to patent application GB 9018805.3, the content of which is incorporated herein by reference.

The correlation between a series of four frames of a 24 frame/s 1:1 format signal and a 30 frame/s, 60 field/s, 3232 pulldown format signal is shown in Figure 49. The first frame A is used to produce the first three fields, with odd field 3 being a phantom field. Frame B produces the next two fields 4 and 5. Frame C produces the next three fields 6 to 8, with even field 8 being a phantom field, and the last frame D in the sequence produces the last two fields 9 and 10.

It is desirable to modify the system described with reference to Figures 1 to 48 so that it can convert a 60 field/s, 30 frame/s 2:1 interlaced video signal to a 60 field/s 3232 pulldown sequence with motion compensation, and this can be achieved in a remarkably simple way by operating the frame recorder 13 and VTR 14 of Figure 2 at one-eighth speed, rather than one-tenth speed (while still maintaining the ten-frame repeat of the output from the standards converter 12) with the period between recorded fields being 9, 7, 9, 7.... field periods. This scheme is illustrated in Figure 50 for five input frames.

As shown in columns A and B of Figure 50, the two fields O/E of each of six input frames 1 to 6 are repeated eight times. Certain of these input fields then used by the progressive scan converter 42 to produce respective progressive scan frames as shown in column D, the progressive scan frames being stored alternately in a pair of frame stores. Certain of the stored progressive scan frames are then used alone or in pairs by the interpolator 48 to produce four fields, each of which is output by the interpolator 48 ten times as shown in column E. A field is then recorded by the frame recorder 13 from every fourth frame of column E, the recorded fields being alternately odd and even with the period between recorded fields being alternately 9/60s and 7/60s, as shown in column F, and these fields are then recorded on the VTR 14. Thus, when the recording is played back at normal speed, the sequence shown in the column G of Figure 50 is produced, with the ten recorded fields in 3232 pulldown format 1 odd, 1 even, 2 odd, 2 even.... 5 even being derived from frame A, repeats 1, 5 and 9; frame 2, repeats 3 and 7; frame 3, repeats 1, 5 and 9; and frame 4, repeats 3 and 7 of the 24 Hz 1:1 format sequence.

As an alternative to the increased recording speed (one-eighth, rather than one-tenth) modification for producing the 3232 pulldown sequence, the progressive scan frames may be written into the frame

recorder 13 at one-tenth speed and then read out in 3232 sequence for supply to the VTR 14. This modification is illustrated in Figure 51. Column E corresponds to column E described above with reference to Figure 50. One repeat of each progressive scan frame is written to the recorder 13, as shown in column F. Odd and even fields of these frames are then read out from the recorder 13 and recorded by the VTR 14 in the following sequence 1 odd, 1 even, 1 odd, 2 even, 2 odd, 3 even, 3 odd, 3 even, 4 odd, 4 even, as shown in column G, to produce fields 1 odd, 2 even, 3 odd, 4 even.... 9 odd, 10 even of the 3232 pulldown sequence, where 3 and 8 are the phantom fields. This modification has an advantage over that described with reference to Figure 50 in that less storage space is used in the recorder 13, and also conversions will require fewer stops and starts of the VTR 14.

It is desirable also to be able to use the system described above to convert 30 Hz 1:1 format film material into 60 field/s 2:1 interlaced HDVS format using motion compensated interpolation. This could be achieved, as shown in Figure 52, by producing the odd fields of converted signal directly from the input frames, and by producing the even fields of the converted signal by motion compensated interpolation between successive fields with equal temporal offsets. However, it is considered that such a scheme would cause uneven spatial response of the field pairs and noise level modulation in the case of noisy source material because the directly produced output fields would contain the source noise, but the interpolated output fields would have reduced noise due to the interpolator action. In order to avoid these problems, a temporal interpolation scheme, as shown in Figure 53A, is adopted, in which the odd fields are temporally interpolated one-quarter of the way between a preceding and a succeeding frame of the source material and the even fields are temporally interpolated three-quarters of the way between those two frames.

Thus, referring to Figure 53B, if a pixel in an odd output field A0 between input frames 1,2 has a position  $(x,y)$  and a motion vector  $(m,n)$ , the value of that pixel is obtained by averaging the value of the pixel (or patch) in input frame 1 at location  $(x-(m/4), y-(n/4))$  with the value of the pixel (or patch) in input frame 2 at location  $(x+(3m/4), y+(3n/4))$ . On the other hand, for the corresponding even output field AE, the output pixel value is obtained by averaging the

value of the pixel (or patch) in input frame 1 at  $(x-(3m/4), y-(3n/4))$  with the value of the pixel (or patch) in input frame 2 at  $(x-(m/4), y+(n/4))$ , as shown in Figure 53C.

5 In order to achieve this, the arrangement described with reference to Figures 1 to 48 is modified or used in the manner which will now be described with reference to Figure 54.

The frames of the 30 Hz 1:1 format film, three of which are shown in column A in Figure 54, are captured by an HD film scanner and converted to video format on tape, as shown in column B. Although odd  
10 "O" and even "E" fields are shown for each frame 1, 2, 3 in column B, it should be remembered that the image data in the two fields of each frame are not temporally offset in the source image. The tape is then played back at one-twentieth speed by the VTR 11 (Figure 2) with the two fields of each frame repeated twenty times, as shown by column C.  
15 Because the two fields of each frame are derived from a progressive format source, there is no temporal offset between them, and therefore the progressive scan converter 42 (Figure 4) is bypassed or operated in a previous field replacement mode, in order to reconstruct the original frames, so that two consecutive frames are available at a time, as  
20 shown in column D. The frames which are input to the direct block matcher 43 and interpolator 48 are thus a direct combination of the respective two fields. In order to produce a pixel in an odd output field, the interpolator temporally interpolates in the ratio  $3/4 : 1/4$  between the previous frame (e.g. 1) and the current frame (e.g. 2).  
25 However, for pixels in an even field, the interpolation is in the ratio  $1/4 : 3/4$  between the previous frame 1 and the current frame 2, and the interpolated frames are each repeated twenty times as shown in column E. Every twentieth frame is written to the frame recorder 13 (Figure 2) as represented in column F of Figure 54 and is then recorded by the  
30 VTR 14, so that when the recording is played back a 60 field/s 2:1 format signal is produced at normal speed, as represented by column G in Figure 54.

It is also desirable to be able to use the system described above to convert 24 Hz 1:1 format film material to 60 field/s 2:1 interlaced  
35 HDVS format. In order to do this, firstly the 24 Hz 1:1 format frames are captured by an HD film scanner and converted to video format on tape. The temporal interpolation scheme which is then used is shown in

Figure 55A. It will be noted that the interpolation sequence repeats after every 4 frames (1 to 4) of the 24 Hz 1:1 format signal and after every 5 frames, or 10 fields (A/odd to E/even), of the 60 field/s 2:1 interlaced HDVS signal.

5 If a pixel in an output field has a motion vector (m,n), then the offsets between the location of that pixel in the output field and the locations in the respective two input frames of the pixels (or patches) used to derive the value of the output pixel are as follows for each of the ten output frames in the series:

10

<u>Output Frame/Field</u>		<u>First Input Frame and Offset</u>	<u>Second Input Frame and Offset</u>
15	A/odd	1 (0,0)	
	A/even	1 (-0.4m,-0.4n)	2 (0.6m,.6n)
	B/odd	1 (-0.8m,-0.8n)	2 (0.2m,0.2n)
	B/even	2 (-0.2m,-0.2n)	3 (0.8m,0.8n)
	C/odd	2 (-0.6m,-0.6n)	3 (0.4m,0.4n)
20	C/even	3 (0,0)	
	D/odd	3 (-0.4m,-0.4n)	4 (0.6m,0.6n)
	D/even	3 (-0.8m,-0.8n)	4 (0.2m,0.2n)
	E/even	4 (-0.2m,-0.2n)	5 (0.8m,0.8n)
	E/even	4 (-0.6m,-0.6n)	5 (0.4m,0.4n)

25

In order to achieve such an interpolation sequence, the system described above with reference to Figures 53 and 54 is modified as follows, referring to Figure 56. In Figure 56, for clarity, the separate fields of frames are not shown. The 24 Hz 1:1 format frames (five of which are shown in column A of Figure 56), after capturing by the HD scanner (column B) are reproduced at one-twentyfifth speed by the VTR 11 (Figure 2) with the two fields of each frame being repeated twenty-five times, as shown in column C. As described with reference to Figure 54, the progressive scan converter 42 is operated in a previous field replacement mode to reconstruct the original frames, and

30

35

so that three frames are available at a time, as shown in column D. The interpolator 48 then performs the necessary interpolation to produce the frames A to E, as shown in column E. It will be noted from the table above, that input frames 1 and 2 are required to produce the two fields of output frame A; input frames 1, 2 and 3 are required for the fields of output frame B; input frames 2 and 3 are required for the fields of output frame C; input frames 3 and 4 are required for the fields of output frame D; and input frames 4 and 5 are required for the fields of output frame E. The frames thus formed by the interpolator are repeated twenty times, as shown in column E. One in every twenty interpolated frames is written to the frame recorder 13, as represented by column F, and recorded on tape by the VTR 14 so that when the tape is played back at normal speed, a 60 field/s 2:1 interlaced HDVS is produced at normal speed, as represented by column G.

The system should also desirably be able to handle 24 Hz 1:1 material provided in 3232 pulldown format. In order to do this, the arrangement of Figure 56 is utilised except that the input tape from the VTR 11 is run at one-twentieth speed, rather than 1/25 speed, and the input frames/fields are repeated twenty times, rather than twenty-five times before the next field is input. The frames of column D are still produced one for every 50 repeated input fields, and therefore the phantom fields of the 3232 pulldown format can be ignored. Accordingly, the system operates somewhat similarly to that described with reference to Figure 54 relating to conversion of 30 Hz 1:1 format film material to a 60 field/s 2:1 interlace format HDVS.

A modification may be made to the arrangement described with reference to Figures 55A and 56 so that it can convert to 30 Hz 1:1 format, rather than 60 field/s 2:1 interface format. With this modification, the output frames are derived from the input frames as shown in Figure 55B. In order to achieve this modification, the only change necessary to make to the scheme of Figure 56 is to modify the action of the interpolator so that, for a pixel in an output frame having a motion vector (m,n), the offsets between the location of that pixel in the output frame and the locations in the respective two input frames of the pixels (or patches) used to derive the value of the output pixel are as follows for each of the five output frames in the series:



<u>Output Frame</u>		<u>First Input Frame</u> <u>and Offset</u>	<u>Second Input Frame</u> <u>and Offset</u>
5	A	1 (0,0)	
	B	1 (-0.8m,-0.8n)	2 (0.2m,0.2n)
	C	2 (-0.6m,-0.6n)	3 (0.4m,0.4n)
	D	3 (-0.4m,-0.4n)	4 (0.6m,0.6n)
	E	4 (-0.2m,-0.2n)	5 (0.8m,0.8n)
10			

It is also desirable that the system described above should be able to convert a 60 field/s 2:1 interlace HDVS format signal to a 30 Hz 1:1 progressive format for use by the EBR 35 (Figure 3) in producing 30 Hz film from the video signal. Such a 30 Hz signal could be produced by compositing each frame from two adjacent fields in the HDVS, but this would have the effect of producing double-imaging in the 30 Hz signal, due to the temporal offset between the two fields making up each frame.

The progressive scan arrangement described above with reference to Figures 4 to 14 can be employed to good effect, to blend different proportions of interframe and intrafield interpolated images depending on the amount of motion locally in the image, and thus provide a motion adapted 30 Hz signal. However, when the source image is noisy or there is an incorrect assessment by the motion adaptive process, the output image will tend to be intrafield interpolated more than is necessary, and thus will lose vertical detail and have more alias components present. In these circumstances, it is possible that the output image quality will be improved by additionally using the motion compensation technique described above with reference to Figures 15 to 48. This will allow two adjacent 30 Hz 1:1 frames to be combined via temporal interpolation and provide cancellation of vertical alias components in static or horizontally moving parts of the image.

The technique using the motion adaptive technique only, is shown in Figures 57 and 58. Referring firstly to Figure 57, and as described above, each frame A, B, C in the output format is produced by different

proportions of three fields (e.g. 1'O', 1'E', 2'O'; 2'O', 2'E', 3'O'; 3'O', 3'E', 4'O') in the input format. Referring now to Figure 58, the 60 field/s 2:1 HDVS format signal (column A) is reproduced at one-eighth normal speed by the VTR 11 (Figure 2), and each frame/field is repeated eight times, as shown in column B. The progressive scan converter 42 (Figure 4) loads fields into its frame stores so that three consecutive fields are available at a time, as shown in column C. The progressive format frames are formed from the triplets of input fields, as shown in column D, and these progressive format frames are each repeated eight times. Every eighth frame (column E) is recorded by the frame recorder 13 and then by the VTR 14 (Figure 2) and thus the recorded signal, when reproduced at normal speed is in 30 Hz 1:1 format (column F). In this motion adaptive mode, the progressive scan converter 42 (Figure 4) is employed, but the motion compensation components 43 to 48 are not. Therefore, the interpolator 48 is set to provide no temporal offset between its input and output frames.

When motion compensation is selected, the operation is as shown in Figures 59 and 60. The upper part of Figure 59 is similar to Figure 57, except that twice as many progressive format frames are produced. The frames so formed then undergo the motion compensation operation to form the output frames which are temporally offset by half a frame period from frames before motion compensation.

Referring in particular to Figure 60, the 60 field/s 2:1 HDVS format signal (column A) is reproduced at one-tenth normal speed by the VTR 11 (Figure 2), and each frame/field is repeated ten times, as shown in column B. As before, the progressive scan converter 42 loads fields into its frames stores so that three consecutive input fields are available at a time, as shown in column C. The progressive scan frames 1'', 2', 2'', 3'.... are then formed from the triplets of input fields, and are loaded into frame stores so that the interpolator 48 has available two consecutive progressive format frames at a time, as shown in column D. Temporally adjacent pairs of these progressive format frames are then used by the interpolator 48 to produce a frame which is interpolated with equal temporal offset between the two input frames. The frames produced by the interpolator 48 are repeated ten times as shown in column E, and every tenth frame (column F) is recorded by the frame recorder 13 and VTR 14 (Figure 2), so that the thus recorded

signal, when reproduced at normal speed, is in 30 Hz 1:1 format (column G).

In the arrangement described above with reference to Figures 1 to 48 for converting 60 field/s 2:1 interlace HDVS to 24 Hz 1:1 film  
5 format, the output frames are either temporally aligned with respective input frames or temporally offset by one half of an input field period (1/120s). In the case of temporal alignment, the output frame is based upon the respective progressive format frame output from the progressive scan converter 42 (Figure 4), whereas in the case of a  
10 temporal offset, each pixel in the output frame is 1/2:1/2 interpolated between pixels or pixel patches in preceding and succeeding progressive format frames output from the progressive scan converter, with spatial offset between the pixels or patches in the source frames and the pixel in the output frame being dependent upon the motion vector supplied by  
15 the processor 47 (Figure 4).

Figure 61A shows the case where there is no temporal offset and a pixel at location (x,y) in the output frame has a motion vector (m,n). This pixel is derived from the pixel at location (x,y) or a patch centred on (x,y) in input frame 1 to the interpolator, and the  
20 motion vector and the content of input frame 2 are not employed.

Figure 61B shows the case where there is a half field period temporal offset and a motion vector of (m,n) = (2,2) for an output pixel at location (x,y). The value of this pixel is derived by equal interpolation between the pixel at (or patch centred on) location (x,y)  
25  $-1/2(m,n) = (x-1, y-1)$  in input frame 1 and the pixel/patch at location  $(x,y) + 1/2(m,n) = (x+1, y+1)$  in input frame 2.

It will be appreciated that the components of the motion vector (m,n) are integers and need not be even integers. Figure 61C shows the case of a motion vector (2,1). The required pixels or patches in the  
30 input frames 1 and 2 are at locations  $(x-1, y-1/2)$  and  $(x+1, y+1/2)$ , respectively, which are half-way between actual pixel positions in the input frames.

In order to acquire the required pixel values from the input frames, an 8 x 8 patch (as described above with reference to Figure 48)  
35 or a 7 x 7 patch as shown in Figures 62A to D is used around the required pixel location, and there is an offset of (0,0) (Figure 62A), (1/2,0) (Figure 62B), (0,1/2) (Figure 62C), or (1/2,1/2) (Figure 62D)

between the centre pixel of the patch and the pixel location (marked "o" in Figures 62A to D) determined by the interpolator. To determine the value of the pixel "o", spatial interpolation coefficients are applied to the 49 pixels in the patch, the sets of coefficients being slightly different for the four cases shown in Figures 62A to D, although the coefficients for the cases of offset  $(1/2, 0)$  and  $(0, 1/2)$  may be symmetrical about the  $x=y$  diagonal of the patch.

A problem which can arise with the arrangement described above is that the magnitude responses for the four different spatial interpolation cases of Figures 62A to D can be different and produce modulation of the picture detail as the different responses are cycled. In order to avoid this problem, the arrangement now described with reference to Figures 63 and 64 may be employed. The locations of the required pixels/patches in frames 1 and 2 are derived in a similar manner to that described with reference to Figure 61 except that an addition offset of  $(-1/4, -1/4)$  is included. Thus, as shown in Figure 63A, for example, for a temporally aligned output frame, a pixel at locations  $(x, y)$  in the output frame is based on an interpolated value of a pixel at location  $(x-1/4, y-1/4)$  in input frame 1, rather than location  $(x, y)$ . When taking into account temporally aligned and temporally offset output frames, and even and odd motion vectors, a required pixel location in an input frame can always be considered to have an offset of  $(-1/4, -1/4)$ ,  $(1/4, -1/4)$ ,  $(-1/4, 1/4)$  or  $(1/4, 1/4)$  from the centre pixel of a  $7 \times 7$  patch from which the value of the required pixel can be calculated by spatial interpolation, as shown in Figures 64A to 64D. It should be noted that these offsets are rotationally symmetrical about the centre pixel of the patch, and accordingly the four sets of spatial interpolation coefficients for the cases shown in Figures 64A to D can also be chosen to be rotationally symmetrical, thus avoiding different magnitude responses and picture detail modulation. It will be appreciated that the above arrangement produces global offset in the output frame of  $(-1/4, -1/4)$  pixels, but this is negligible.

From the above it will be appreciated that, with conversion from 60 field/s 2:1 interlace HDVS to 24 Hz 1:1 film format, every other output frame from the interpolator 48 is produced from one of the progressive frames input to the interpolator, and the alternate output

frames are produced by motion compensation between two progressive frames input to the interpolator. This can result in (a) perspective changes not being satisfactorily merged, (b) alias effects when the adaptive progressive scan conversion fails due to noise, and (c) noise level modulation when the input image is noisy. As regards point (b), when adaptive progressive scan fails due to noise, the progressive scan frames are produced by intrafield interpolation. If such a frame is directly output by the interpolator 48, stationary images would appear heavily aliased.

10        In order to reduce these problems, in the case of a temporally aligned output frame from the interpolator 48, the frame is produced by equal summing of the two respective input frames to the interpolator, but with the motion vector being used to determine the spatial offset between the respective pixel in the output frame and the pixel/patch to be used in frame 2, without there being any spatial offset dependent upon the motion vector between the pixel in the output frame and the pixel/patch to be used in frame 1. This scheme is illustrated in Figure 65, in combination with the quarter pixel offset scheme described above with reference to Figures 63 and 64. In the example given, a pixel at location  $(x,y)$  in the output frame has a motion vector  $(2,1)$ . Accordingly, the notional source pixel to be used from frame 1 has a location  $(x-1/4, y-1/4)$ , and therefore the  $7 \times 7$  patch centred on location  $(x,y)$  is used with the set of spatial interpolation coefficients for an offset of  $(-1/4, -1/4)$  (Figure 64A). On the other hand, the notional source pixel to be used from frame 2 has a location  $(x+7/4, y+3/4)$ , and therefore the  $7 \times 7$  patch to be used is centred on location  $(x+2, y+1)$  with the spatial interpolation coefficient set also for an offset  $(-1/4, -1/4)$ .

20        By producing each pixel in the output frame by equal summing from two input frames, whether or not there is a temporal offset between the output frame and the input frames, alias is removed, because the frame 2 alias will always be in antiphase to the frame 1 alias as long as the interfield motion is an exact multiple of lines. As synthesised lines are mixed with non-synthesised lines in this scheme, an improved vertical response is also produced. A further advantage is that if there is noise in the input image, the noise will not be modulated, unlike the case where every other output frame is derived from only one

of the input frames.

There now follows a description of a particular system using the apparatus described above which is used primarily for transfer from 24 Hz film to 24 Hz film permitting HDVS post production. Due to the complexity of motion compensated interpolation processing, the equipment therefor tends to be large and expensive. In addition, there is always a risk that processing artifacts may be introduced into the video signal. For these reasons, it is desirable that only a single stage of motion compensated processing should be used between the source(s) and primary distribution path, if at all possible.

Referring to Figure 66, the first source is 24 Hz film material 100, and the primary distribution medium is also 24 Hz film material 102. The source film 100 is read by a high definition film scanner 104, which incorporates a 3232 pulldown system 106 to provide a high definition 60 Hz 3232 format signal which is recorded by VTR 108.

The second source is a camera 110 which provides a 60 field/s 2:1 interlace HDVS signal to a VTR 112. The VTR 112 can reproduce the HDVS signal for input to a standards converter 114 which converts the 60 field/s 2:1 HDVS signal to a 60 Hz 3232 format signal which is recorded on VTR 116. The standards converter is therefore as described above with reference to Figures 49 to 51. Upon reproduction of the 60 field/s 3232 format signals by the VTRs 108, 116, these signals can be integrated in the HDVS post production system 118 to produce an output signal in 60 Hz 3232 format which is recorded by VTR 120. For further information concerning post-production with a 3232 pulldown format signal, reference is directed to patent application GB9018805.3, the content of which is incorporated herein by reference. Upon reproduction of the signal by the VTR 120, the primary path is via a drop field device 122 which converts the signal to 24 Hz 1:1 format, which is then used by the electron beam recorder (EBR) 124 to generate the film 102. Secondary distribution paths are provided by a VTR 126 which can record the output signal from the drop field device 122, and the signal on reproduction is then fed to a 625 lines converter 128 which produces a 50 field/s 2:1 interlace 625 line signal and/or to a high definition 50 Hz 2:1/1:1 converter 130 which produces a 50 Hz 1:1 high definition video signal. It will be appreciated that there will be a 4% speed error in the signals output from the converters 128 and

130, due to the input frame frequency being 24 Hz, rather than 25 Hz.

Further secondary output paths are provided from the VTR 120 via an HDVS to NTSC down converter 132, which produces an NTSC 59.94 Hz 2:1 signal. Furthermore, there is a direct output of the 60 Hz 3232 format  
5 signal on line 134. Also, a standards converter 136 is included for converting the 60 Hz 3232 format signal from the VTR 120 to a 60 field/s 2:1 interlace HDVS signal. It will therefore be appreciated that the standards converter 136 is as described above with reference to Figure 56. The output of the standards converter 136 may also be  
10 fed to an HDVS to NTSC down converter 138 which produces an NTSC 59.94 Hz 2:1 signal which will be of more acceptable quality than the signal provided by the converter 132, due to proper removal of the phantom fields by the standards converter 136.

The above arrangement has the following features and advantages.  
15 Firstly, there is a single stage of motion compensated interpolation, in standards converter 114, between all acquisition media (film 100 and camera 110) and the primary distribution path on 24 Hz film. Secondly, the post production system 118 allows integration of 60 field/s 2:1 HDVS originated material from the camera 110 or other such source with  
20 material originated from 24 Hz film. Thirdly, there is provided a secondary distribution route as a 60 Hz 3232 format HDVS signal on line 134 which may well be acceptable for many applications. This obviates the requirement to post produce video and film release versions of material separately. Fourthly, there is an additional or alternative  
25 secondary distribution route by way of the 60 field/s HDVS signal from the standards converter 136. Lastly, means are provided for converting the video signal to conventional definition NTSC (for use in U.S.A. and Japan) and for converting to both high definition and 625 lines signals at 25 field/s with 2:1 interlace.

30 There now follows a description of a particular system employing the apparatus described above which is used primarily for transfer from 24 Hz film to 60 field/s 2:1 interlaced HDVS permitting HDVS post production. Again, due to the complexity of motion compensated interpolation processing, the equipment therefor tends to be large and  
35 expensive. In addition, there is always a risk that processing artifacts may be introduced into the video signal. For these reasons it is important that only a single stage of motion compensated

processing should be used between the source(s) and primary distribution path, if at all possible.

Referring to Figure 67, the first source is 24 Hz film material 140, and the primary distribution is by way of a 60 field/s 2:1 HDVS signal on line 142. The source film 140 is read by a high definition film scanner 144, which provides a 24 Hz 1:1 signal to a VTR 146. The signal reproduced by the VTR 146 is fed to a standards converter 148 which converts the 24 Hz 1:1 format signal to 60 field/s 2:1 interlaced format signal, which is recorded by a VTR 150. It will therefore be appreciated that the standards converter 148 is as described above with reference to Figure 55.

Second and third sources are in the form of a camera 152 and a VTR 154, each of which produce 60 field/s 2:1 interlaced video signals. The signals from the VTR 150, camera 152 and VTR 154 can be integrated in the HDVS post production system 156 to produce a 60 field/s 2:1 interlace HDVS signal, which can be recorded by the VTR 158. The primary output path is directly from the VTR 158 on line 148, which carries the reproduced 60 field/s 2:1 interlaced HDVS.

Secondary distribution paths are provided from the VTR 158 via an HDVS to NTSC down converter 160 which outputs a standard NTSC signal and via an HDVS to 1250/50 high definition converter 162 which provides a 50 frame/s 1:1 video signal. A 625 lines 50 field per second 2:1 interlaced signal can either be provided by a 1250/50 to 625 lines converter 164 connected to the output of the converter 162, or by an HDVS to 625/50 down converter 166 receiving the 60 field/s 2:1 interlaced HDVS signal from the VTR 158.

A further secondary distribution path is provided by a standards converter 168 which converts the 60 field/s 2:1 interlaced HDVS signal from the VTR 158 to 24 Hz 1:1 format which is recorded by the VTR 170. The standards converter 168 is therefore as described above with reference to Figures 1 to 48. The reproduced signal from the VTR 170 supplies an EBR 172 to produce 24 Hz 1:1 film 174.

A further secondary distribution path is via a standards converter 176 which receives the 60 field/s 2:1 interlaced HDVS from the VTR 158 and provides a 30 Hz 1:1 format signal to a VTR 178. The converter 176 is therefore as described above with reference to Figures 57 to 60. The VTR 178 reproduces the 30 Hz 1:1 signal to an EBR 180,



which produces a 30 Hz 1:1 film 182.

The system described with reference to Figure 67 has the following features and advantages. Firstly, there is a single stage of motion compensated interpolation, in the standards converter 148, 5 between all of the acquisition media and the primary distribution path on line 142. Secondly, the system allows post production integration of 60 field/s 2:1 interlaced HDVS originated material with 24 Hz film material, and the camera 152 may be used live into the post production chain. A secondary distribution path is provided to output 30 Hz film, 10 and in this case the standards converter 176 may be used without motion compensated interpolation processing, using only motion adaptive interpolation, as described above with reference to Figures 57 and 58. A means is provided of outputting 24 Hz film, but this does entail a second stage of motion compensated interpolation. The system also 15 provides means for down converting to conventional definition NTSC (for U.S.A. and Japan) and also means for converting to both high definition and 625 lines format at 25 Hz frame rate.

There now follows a description of a particular system employing the apparatus described above which is used primarily for transfer from 20 30 Hz film to 30 Hz film permitting HDVS post production. Again, due to the complexity of motion compensated interpolation processing, the equipment therefor tends to be large and expensive. In addition, there is always a risk that processing artifacts may be introduced into the video signal. For these reasons it is important that the number of 25 stages of motion compensated processing between the source(s) and primary distribution path should be as few as possible.

Referring to Figure 68, the first source is 30 Hz film material 190, and the primary output is 30 Hz film 192. The source film 190 is read by a high definition film scanner 194, which provides a 30 Hz 1:1 30 signal to a VTR 196.

Second and third sources are in the form of a camera 198 and a VTR 200, each of which produce 30 Hz 1:1 video signals. Fourth and fifth sources are in the form of a camera 202 and a VTR 204, each of which provide a 60 field/s 2:1 interlace HDVS signal to a standards 35 converter 206 which converts the signals to 30 Hz 1:1 format with motion adaptive interpolation, but not motion compensated interpolation. The converter 206 is therefore of the form described

above with reference to Figures 57 and 58.

The 30 Hz 1:1 format signals from the VTR 196, camera 198, VTR 200 and standards converter 206 can be integrated in the HDVS post production system 208 to produce an output 30 Hz 1:1 video signals, which can be recorded by the VTR 210. The primary output path is from the VTR 210

to an EBR 212 which produces the output 30 Hz 1:1 film 192.

Secondary distribution paths are provided from the VTR 210 via a standards converter 214 to a VTR 216. The converter 214 converts the input 30 Hz 1:1 video signals to 60 field/s 2:1 interlace format, and is therefore of the form described above with reference to Figure 54. The signal reproduced by the VTR 216 can therefore be directly output on the line 218 as a 60 Hz 2:1 interlace HDVS signal, and can also be converted by converters 220, 222, 224 respectively to: NTSC format; 1250/50 format; and 625 lines 50 field/s 2:1 interlace format. Furthermore, the 60 field/s 2:1 interlace HDVS reproduced by the VTR 216 can be converted by a standards converter 226, of the type described above with reference to Figures 1 to 48, to 24 Hz 1:1 format for producing, via a VTR 228 and EBR 230, 24 Hz 1:1 film 232. The 24 Hz 1:1 signal reproduced by the VTR 228 may also be line rate converted by a converter 234 to produce a pseudo 25 Hz frame rate video signal.

The system described with reference to Figure 68 has the following features and advantages. Firstly, there are no stages of motion compensated interpolation between any of the acquisition media 190, 198, 200, 202 and 204 and the primary distribution path on 30 Hz 1:1 film 192. Secondly, the system allows post production integration of video originated material with 30 Hz film material, and the camera 198 may be used live in the post production chain. Secondary distribution paths are provided in 60 field/s 2:1 interlaced format and 59.94 Hz NTSC format with acceptable motion characteristics and without complex motion compensated interpolation. Motion portrayal in the 60 field/s 2:1 interlaced HDVS signal and NTSC signal is enhanced by the motion compensated progressive to interlace conversion by the standards converter 214. A means is provided of outputting 24 Hz film, but this does entail a second stage of motion compensated interpolation. The system also provides means for down converting to conventional definition NTSC (for U.S.A. and Japan) and also means for converting to

both high definition and 625 lines format at 25 Hz frame rate. The standards converter 206 permits standard HDVS 2:1 interlaced cameras 202 and VTRs 204 to be used for video acquisition, but their outputs are converted to 30 Hz 1:1 format by a motion adaptive process, and therefore the vertical resolution of moving images will be more limited than in the case of the 30 Hz 1:1 camera 198 and VTR 200. The system of Figure 68 requires the post production chain to process the images in progressive scan format and in this connection reference is directed to patent application GB 9018805.3, the content of which is incorporated herein by reference.

There now follows a description of a particular system employing the apparatus described above which is used primarily for transfer from 30 Hz or 60 Hz film to 60 field/s 2:1 interlaced HDVS signals permitting HDVS post production. Again, due to the complexity of motion compensated interpolation processing, the equipment therefor tends to be large and expensive. In addition, there is always a risk that processing artifacts may be introduced into the video signal. For these reasons it is important that not more than one stage of motion compensated processing should be used between the source(s) and primary distribution path, if at all possible.

Referring to Figure 69, the first source is 30 Hz film material 240 or 242; the second source is 60 Hz film material 244; the third and fourth sources are a 60 field/s 2:1 interlace HDVS format camera 246 and VTR 248; and the primary distribution is by way of a 60 field/s 2:1 HDVS on line 250.

The 30 Hz source film 242 is read by a high definition film scanner 258, which provides a 30 Hz 1:1 signal to a VTR 260. Upon reproduction of the signal by the VTR 260, it is converted, not necessarily at real-time rate, to 60 field/s 2:1 interlace HDVS format by a standards converter 262, of the type described above with reference to Figure 54, the converted signal being recorded by a VTR 264. As an alternative, in the case that real-time conversion from 30 Hz 1:1 to 60 field/s 2:1 interlace format becomes possible, the 30 Hz source film 240 may be read by a high definition film scanner 252, which provides a 30 Hz 1:1 signal to a VTR 254. Upon reproduction of the signal by the VTR 254, it is real-time converted to 60 field/s 2:1 interlace format HDVS by a converter 256. In the case of the 60 Hz

film 244, it is read by a high definition film scanner 266, which incorporates a device 268 for pull down on every field, so that a 60 Hz 2:1 interlace format HDVS signal is produced, which is recorded by a VTR 270.

5       The 60 Hz 2:1 interlace format video signals from the converter 256, VTR 264, VTR 270, camera 246 and/or VTR 248 can be integrated in the HDVS post production system 272 to produce a 60 field/s 2:1 interlace HDVS signal, which can be recorded by the VTR 274. The primary output path is directly from the VTR 274 on line 250, which  
10 carries the reproduced 60 field/s 2:1 interlaced HDVS signal.

      Secondary distribution paths are provided from the VTR 274 via an HDVS to NTSC down converter 276 which outputs a standard NTSC signal and via an HDVS to 1250/50 high definition converter 278 which provides  
15 a 50 frame/s 1:1 video signal or 50 field/s 2:1 interlace video signal. A 625 lines 50 field per second 2:1 interlaced signal can either be provided by a 1250/50 to 625 lines converter 280 connected to the output of the converter 278, or by an HDVS to 625/50 down converter 282 receiving the 60 field/s 2:1 interlaced signal from the VTR 274.

      A further secondary distribution path is provided by a standards  
20 converter 284 which converts the 60 field/s 2:1 interlaced HDVS signal from the VTR 274 to 24 Hz 1:1 format which is recorded by the VTR 286. The standards converter 284 is therefore as described above with reference to Figures 1 to 48. The reproduced signal from the VTR 286 supplies an EBR 288 to produce 24 Hz 1:1 film 290.

25       A further secondary distribution path is via a standards converter 292 which receives the 60 field/s 2:1 interlaced HDVS from the VTR 274 and provides a 30 Hz 1:1 format signal to a VTR 294. The converter 292 is therefore as described above with reference to Figures 57 to 60. The VTR 294 reproduces the 30 Hz 1:1 signal to an EBR 296,  
30 which produces a 30 Hz 1:1 film 298.

      The system described with reference to Figure 69 has the following features and advantages. Firstly, there is a single stage of motion compensated interpolation, in the standards converter 262 for the 30 Hz film 258, between all of the acquisition media and the  
35 primary distribution path on line 250. Secondly, the system allows post production integration of 60 field/s 2:1 interlaced HDVS originated material with 30 Hz and 60 Hz film material, and the camera

246 may be used live in the post production chain. A secondary distribution path is provided to output 30 Hz film, and in this case the standards converter 284 may be used without motion compensated interpolation processing, using only motion adaptive interpolation, as  
5 described above with reference to Figures 57 and 58. A means is provided of outputting 24 Hz film, but this does entail a second stage of motion compensated interpolation. The system also provides means for down converting to conventional definition NTSC (for U.S.A. and Japan) and also means for converting to both high definition and 625  
10 lines format at 25 Hz frame rate.

There now follows a description of a particular system employing the apparatus described above which is used primarily for transfer from 30 Hz film to 24 Hz film permitting HDVS post production. Again, due to the complexity of motion compensated interpolation processing, the  
15 equipment therefor tends to be large and expensive. In addition, there is always a risk that processing artifacts may be introduced into the video signal. For these reasons it is important that the number of stages of motion compensated processing between the source(s) and primary distribution path should be as few as possible.

20 Referring to Figure 70, the first source is 30 Hz film material 300, and the primary output is 24 Hz film 302. The source film 300 is read by a high definition film scanner 304, which provides a 30 Hz 1:1 signal to a VTR 306.

Second and third sources are in the form of a camera 308 and a  
25 VTR 310, each of which produce 30 Hz 1:1 video signals. Fourth and fifth sources are in the form of a camera 312 and a VTR 314, each of which provide a 60 field/s 2:1 interlace HDVS signal to a standards converter 316 which converts the signals to 30 Hz 1:1 format with motion adaptive interpolation, but not necessarily motion compensated  
30 interpolation. The converter 316 is therefore of the form described above with reference to Figures 57 and 58.

The 30 Hz 1:1 format signals from the VTR 306, camera 308, VTR 310 and standards converter 316 can be integrated in the HDVS post production system 318 to produce an output 30 Hz 1:1 video signals,  
35 which can be recorded by the VTR 320. The primary output path is from the VTR 320 via a standard converter 322 which converts to 24 Hz 1:1 format and a VTR 324 to an EBR 326 which produces the output 24 Hz 1:1

film 302.

A secondary distribution path is provided from the VTR 210 to an EBR 328 which produces 30 Hz 1:1 format film 330. Further secondary distribution paths are via a standards converter 332 to a VTR 334. The converter 332 converts the input 30 Hz 1:1 video signal to 60 field/s 2:1 interlace HDVS format, and is therefore of the form described above with reference to Figure 54. The signal reproduced by the VTR 334 can therefore be directly output on the line 336 as a 60 Hz 2:1 interlace HDVS signal, and can also be converted by converters 338, 340, 342 respectively to: NTSC format; 1250/50 format; and 625 lines 50 field/s 2:1 interlace format.

The system described with reference to Figure 70 has the following features and advantages. Firstly, there is only one stage of motion compensated interpolation (in converter 332) between any of the acquisition media 300, 308, 310, 312 and 314 and the primary distribution path on 24 Hz 1:1 film 302, and there is no motion compensated interpolation between the inputs and the output 30 Hz 1:1 film. Secondly, the system allows post production integration of video originated material with 30 Hz film material, and the camera 308 may be used live in the post production chain. Secondary distribution paths are provided in 60 field/s 2:1 interlaced format and 59.94 Hz NTSC format with acceptable motion characteristics and without complex motion compensated interpolation. Motion portrayal in the 60 field/s 2:1 interlaced video signal and NTSC signal is enhanced by the motion compensated progressive to interlace conversion by the standards converter 332. The system also provides means for down converting to conventional definition NTSC (for U.S.A. and Japan) and also means for converting to both high definition and 625 lines format at 25 Hz frame rate. The standards converter 316 permits standard HDVS 2:1 interlaced cameras 312 and VTRs 314 to be used for video acquisition, but their outputs are converted to 30 Hz 1:1 format by a motion adaptive process, and therefore the vertical resolution of moving images will be more limited than in the case of the 30 Hz 1:1 camera 198 and VTR 200. The system of Figure 70 requires the post production chain to process the images in progressive scan format and in this connection reference is directed to patent application GB 9018805.3, the content of which is incorporated herein by reference.

In the arrangement described above with reference to Figure 2, the VTR 11 plays at one-eighth speed into the standards converter 12, and the standards converter provides ten repeats of each output frame. The frame recorder 13 stores one in every ten of the repeated output frames until it is full, and the stored frames are then output at normal speed to the VTR 14 which records at normal speed. The material is therefore converted in segments, entailing starting, stopping and cuing of both of the VTRs 11,14. In order to convert one hour of source material using a frame recorder 13 with a capacity of 256 frames, it is necessary to start, stop and cue each of the recorders 338 times, and it will be realised that this can cause considerable wear of both recorders. Furthermore, the operations of alternately reading from the VTR 11 and then recording on the VTR 14, with cuing of both recorders results in the conversion process being slow. Indeed, in the example given above, although the standards converter processes at one-eighth speed, the conversion of one hour of material would take not 8 hours, but almost 11 hours. With smaller capacity frame recorders 13, the wasted time would be increased.

The arrangement of Figure 71 will now be described, which is designed to increase the conversion rate to the maximum possible. Instead of one frame recorder 13 of 256 frame capacity, two frame recorders 13A, 13B are provided, each of 128 frame capacity, each receiving the output of the standards converter 12, and each controlled by the system controller 15. The outputs of the frame recorders 13A, 13B are fed to a 2:1 digital multiplexer 13C, which selects the output from one or the other of the frame recorders 13 under control of the system controller 15 and supplies the selected signal to the VTR 14. With this modification, the source VTR 11 is operated non-stop, and every tenth frame in a series of 1280 frames output by the standards converter are recorded, alternately 128 frames by one frame recorder 13A and 128 frames by the other frame recorder 13B. When one frame recorder is recording, the other frame recorder has time to play back its stored 128 frames to the VTR 14. Thus, in the conversion of 1 hour of material, the VTR 11 starts and stops once, and the VTR 14 starts and stops 675 times, with sufficient time between stops and starts for the VTR 14 to be used. Accordingly, the conversion time for 1 hour of material is 8 hours, as limited by the processing rate of the standards

converter.

As an alternative to the arrangement of Figure 71, a system as shown in Figure 2 may be used, but in which a frame recorder 13 is used which permits simultaneous recording and playback and which is operable with a cyclic frame addressing scheme. Thus, frames from the standards converter 12 can be stored at the relatively low rate dictated by the converter 12, and then, when the frame recorder is nearly full, the frames can be played back to the VTR 14 at the relatively high rate required by the VTR 14 (while the frame recorder is still being fed by the converter 12) leaving space in the frame recorder to be overwritten by further input frames. The use of this type of frame recorder reduces the memory requirement, as compared with the arrangement of Figure 71, and also obviates the need for the multiplexer 13C.

The arrangement of Figure 72 will now be described, which is designed to reduce the amount of wear of the VTRs 11, 14. The arrangement of Figure 72 is similar to that of Figure 2, except that the path between the VTR 11 and the VTR 14 is either via the standards converter 12 when switches 11A, 14A under control of the system controller 15 are each in position "1", or via the frame recorder 13 when the switches 11A, 14A are each in position "2". Operation is in two phases: phase 1 when the switches are in position 1 and then phase 2 when the switches are in position 2.

In the following description it is considered that the tapes on the VTRs 11, 14 have a number of sequential positions or slots for frames, which are numbered 0 to 86399, plus some spare, in the case of recording 1 hour of frames in 24 Hz 1:1 format, and that the frames of the 1 hour of material to be recorded in 24 Hz 1:1 format are numbered sequentially 0 to 86399. It is also assumed that the frame recorder 13 has a capacity C of 253 frames.

In phase 1, the VTR 11 and VTR 14 are operated intermittently and simultaneously. When VTR 11 is playing, the standards converter 12 produces a series of frames each repeated R times where  $R = 10$ . The VTR 11 is operated so that  $RC(R + 1) = 27830$  frames (including the repeats) are produced by the standards converter 12, and one in every  $R (= 10)$  of these frames is recorded by the VTR 14 at normal speed starting at frame slot 0, so that frames 0 to 2782 of the material are



recorded in frame slots 0, 10, 20.... 27820 on the tape on the VTR 14, with the intermediate frame slots left blank. The VTR is then cued back to start reading again at a frame slot offset from the previous starting frame slot by  $C(R + 1) = 2783$  frames, so that for this second  
5 recording run the starting slot is 2783. A further 27830 frames (including repeats) are produced by the converter and one in ten is recorded, thus at frame slots 2783, 2793, 2803..... 30603 of the tape on the VTR. Re-cuing and frame production and recording continues like  
10 this, and the table below gives examples of the numbers of frames and the frame slots at which they are recorded, after the VTR's 11, 14 have been started and stopped  $1 + \text{INT}(86399/2783)$  ( $=32$ ) times and all 86400 frames have been recorded by the VTR 14.

	Pass No.	First Frame No.		Last Frame No.	
		Slot		Slot	
5	0	0	1	---	2782
		0	10	---	27820
	1	2783	2784	---	5565
		2783	2793	---	30603
10	2	5566	5567	---	8348
		5566	5576	---	33386
	3	8349	8350	---	11131
		8349	8359	---	36169
15	4	11132	11133	---	13914
		11132	11142	---	38952
20	5	13915	13916	---	16697
		13915	13925	---	41735
	6	16698	16699	---	19480
		16698	16708	---	44518
25	7	19481	19482	---	22263
		19481	19491	---	47301
	8	22264	22265	---	25046
		22264	22274	---	50084
30	9	25047	25048	---	27829
		25047	25057	---	52867
35	10	27830	27831	---	30612
		27830	27840	---	55650
	:	:	:	---	:
		:	:	---	:
40	31	86273	86274	---	86399
		86273	86283	---	87533

45 If the above table is analyzed, it will be noted that any particular frame, having frame number F, is recorded during pass number  $P = \text{INT}(F/C(R+1))$ , and that it is recorded at slot number  $S = RF - CP(R^2-1)$ . It should be also noted that nearer the beginning and end of the recorded tape, not every frame slot is used. For example, between slots 0 and 2783, nine out of ten slots are left blank, and between slots 2783 and 5566 eight out of ten are blank.

50 In the second phase, the tape recorded on the VTR 14 during phase

1 is loaded onto the VTR 11; a fresh tape is loaded onto the VTR 14; and the switches 11A, 14A are moved to position "2". The tapes on the VTRs 11, 14 are then run continuously at normal speed and rewind repeatedly until the tapes have made  $R+1$  ( $=11$ ) passes through the VTRs  
5 11, 14. During each pass, selected frames from the VTR 11 are stored in the frame recorder 13 until it has reached its capacity and the stored frames are then output to and recorded by the VTR 14; and this is repeated until the end of the recorded tape is reached. For each pass, there is a different offset between the slot numbers of the tapes  
10 on the two VTRs 11, 14.

More specifically, for each pass  $P$  ( $P=0$  to  $R$ ), the VTR 14 is started after the VTR 11 so that there is an offset between the slot number  $S_1$  of the tape on the VTR 14 and the slot number  $S_0$  of the tape on the VTR 11 of  $S_0 - S_1 = C(PR + R - P)$ . Starting at slot number  $S_0 = PCR$  on  
15 the VTR 11, every  $R$ th ( $=10$ th) frame is stored in the frame recorder 13 until it is full, i.e. has stored  $C$  frames. With both VTRs 11, 14 running, the stored frames are output from the frame recorder 13 at normal speed and recorded by the VTR 14. This is then repeated so that every next  $R$ th frame from the VTR 11 is stored in the frame recorder  
20 13, and so on. A specific example of this is shown in the table below, where  $F_0$  is the original frame number of a frame input to the frame recorder and  $F_1$  is the original frame number at a frame output from the frame recorder.

	<u>Pass 0</u>					<u>Pass 1</u>			
	<u>S<sub>0</sub></u>	<u>E<sub>0</sub></u>	<u>E<sub>1</sub></u>	<u>S<sub>1</sub></u>		<u>S<sub>0</sub></u>	<u>E<sub>0</sub></u>	<u>E<sub>1</sub></u>	<u>S<sub>1</sub></u>
5	0	0		-		2530	253		-
	10	1		-		2540	254		-
	20	2		-		2550	255		-
	:	:		-		:	:		-
	2520	252		-		5050	505		-
10									
	2530		0	0		5060		253	253
	2531		1	1		5061		254	254
	:		:	:		:	:	:	:
15	2782		252	252		5312		505	505
	2783	2783		253		5313	3036		506
	2793	2784		263		5323	3037		507
20	:	:		:		:	:		:
	5303	3035		2773		7833	3288		3026
	5313		2783	2783		7843		3036	3036
25	5314		2784	2784		7844		3037	3037
	:		:	:		:		:	:
	5565		3035	3035		8095		3288	3288
30	:	:		:		:	:		:
	:	:		:		:	:		:

	<u>Pass 10</u>				<u>Pass P (General case)</u>			
	$S_0$	$E_0$	$E_1$	$S_1$	$S_0$	$E_0$	$E_1$	$S_1$
5	25300	2530		-	PCR	CP		
	25310	2531		-	+10	+1		
	:	:		-	:	:		
	27820	2782		-	+10	+1		
10	27830		2530	2530	+10		CP	CP
	27831		2531	2531	+1		+1	+1
	:		:	:	:		:	:
	28082		2782	2782	+1		+1	+1
15	28083	5313		2783	+1	C(P+R+1)		+1
	28093	5314		2793	+10	+1		+10
	:	:		:	:	:		:
20	30603	5565		5303	+10	+1		+10
	30613		5313	5313	+10		C(P+R+1)	+10
	30614		5314	5314	+1		+1	+1
25	:		:	:	:		:	:
	30865		5565	5565	+1		+1	+1
30	:	:		:	+1	C(P+2R+2)		+1
	:	:		:	+10	+1		+10
					:	:		:

From an analysis of the above tables it will be appreciated that the original frames are recorded stage by stage at the appropriate frame slots on the tape on the VTR 14, so that after the R+1 passes all of the original frames are recorded in the correct order on the tape.

For the above system to operate satisfactorily without frames on the intermediate tape being overwritten, it is necessary that (C modulo R) is non-zero, and that (C modulo R) and R do not have a common factor. Values of C and R may be chosen which do not satisfy these conditions, but it is then necessary to employ a more complicated system to determine the starting slots of the intermediate tape in phase 1, and the offsets  $S_0$ - $S_1$  in phase 2, and possibly to make more than (R + 1) passes of the tapes in phase 2.

During phase 1 of the above procedure, the number of starts and stops of each VTR 11, 14 is 1+ the integral part of ((highest frame number)/C(R+1)) which is 1+INT(86399/2783) = 32 in the example given, and thus 64 total for both VTRs 11, 14. In phase 2, there are (R+1)

starts and stops of each VTR, giving 22 for both recorders. Accordingly, the total number of starts and stops is 86 over both phases, which compares favourably with the figure of 676 for the arrangement of Figure 2 without this modification.

5        In the arrangement described above, it is necessary for the VTR  
11 to be able to provide a slow motion output, for example at one-  
eighth speed in the case of conversion from 60 field/s 2:1 interlace  
format to 24 Hz 1:1 format. If a versatile system is to be provided  
capable of converting between a variety of different formats, then the  
10 standards converter 13 requires a variety of input speeds, e.g. 1/8,  
1/10, 1/20th. In order to obviate the need for a VTR capable of a  
variety of playback speeds, the arrangement of Figure 73 may be used.  
In this arrangement, a frame recorder 11A is placed in the path between  
the VTR 11 and the standards converter 12 under control of the system  
15 controller 15. The VTR 11 is operated at normal speed, and a series of  
the output frames (e.g. 256) are stored in the frame recorder 11A.  
When full, the frame recorder outputs each stored frame to the  
standards converter 12 the required number of times, for example ten  
times to simulate one-tenth speed, meanwhile the VTR 11 is cued ready  
20 to supply the next 256 frames to the frame recorder 11A once all of the  
stored frames have been output to the standards converter.

With the arrangement of Figure 73, the standards converter cannot  
be operated continuously, because time has to be allowed for frame  
recorder 11A to be refreshed. In order to avoid this problem, a  
25 special cyclically addressable frame store as described above may be  
used, or alternatively the modification as shown in Figure 74 may be  
made. In Figure 74, a pair of frame recorders 11B, 11C are provided in  
parallel, and each of which can output via a 2:1 digital multiplexer  
11D (controlled by the system controller 15) to the standards converter  
30 12. Thus, when either frame recorder 11B, 11C is outputting to the  
standards converter 12, there is time for the next series of frames to  
be stored in the other frame recorder. Accordingly, continuous output  
to, and operation of, the standards converter 12 is permitted.

It will be appreciated that the modifications described above  
35 with reference to Figure 71 and Figures 73 or 74 may be combined to  
permit continuous conversion at less than real-time speed without the  
need for slow-motion replay.

In a motion compensated interpolating system as described above with reference to Figures 1 to 48, only a small number of motion vectors can be tested on a pixel-by-pixel basis. For optimum operation of the system it is important that the best vectors are pre-selected for testing by the motion vector selector 46. Techniques using global motion vectors only have proved to be good for many types of picture and techniques using only locally derived motion vectors have proved good for certain material. Neither is good for all material.

In order to improve on the system described above, a concept will now be described of dividing a picture up into large subdivisions and then calculating which motion vectors are detected most frequently in each of those subdivisions. This technique is an intermediate technique which combines good points from both of the above mentioned approaches.

The technique described above includes the process of counting the most frequently detected motion vectors within a given field and making these motion vectors available for use throughout the picture. Calculating the most frequent vectors over the whole picture area and then applying them in some overall 'vector reduction' strategy has the advantage of providing 'likely' vectors in areas where results obtained from immediately surrounding pixels are inconclusive. The technique described above with reference to Figures 1 to 48 includes the process of 'growing' which is a technique of two-dimensional area summing of correlation surfaces derived from block matching with those derived from adjacent areas of the picture (as described particularly with reference to Figure 21) to enlarge the area over which the match is performed if the nature of the original surface does not permit a good vector to be calculated. Vector reduction as described before considers pictures in progressively larger blocks in order to discover satisfactory vectors to be applied to pixels within that region. These blocks start with a single 'search area' of, for example, 32 x 24 pixels which can then be 'grown' in a variety of ways up to a maximum of, for example, 160 x 72 pixels. Thereafter global vectors are derived from the entire picture area of, for example, 1920 x 1035 pixels.

The advantage of considering a range of different block sizes when determining vectors is that the area which just encompasses a

moving object over two field intervals is nearly optimum for discovering that motion vector. Thus small blocks favour small objects and large blocks favour large objects, such as a panning background.

5 In the strategy described with reference to Figures 1 to 48, no area larger than  $160 \times 72$  and smaller than  $1920 \times 1035$  is considered. However, by subdividing the picture into a number of regular adjoining or overlapping areas and deriving an intermediate vector or vectors for each of those areas in turn it is possible to favour the vectors  
10 pertaining to larger objects whose features make reliable detection of vectors over their whole surface by methods described before difficult, but which are still too small relative to the overall picture to be significant in a list of global vectors.

The subdivision of the picture area may be done in a number of ways. For example, the picture area 350 may be divided into a regular  
15 array, for example a  $4 \times 3$  array, of intermediate areas 352, as shown in Figure 76, or a non-regular array, for example a  $3 \times 3$  array, of intermediate areas, as shown in Figure 78, with, for example, the centre area 354 smaller than the rest so that the intermediate vector or vectors for that area are more localised than for the other areas.  
20 The intermediate areas may adjoin, but be distinct, as in Figures 76 and 78, or they may overlap as shown by the example areas in Figure 77. Thus, in Figure 77, the motion vectors available for an output pixel P which lies both in area 356 and 358 are: the global vector(s), the intermediate vector(s) for area 356; the intermediate vector(s) for  
25 area 358; and the local vector(s) for the pixel P. Alternatively, as shown in Figure 75, the intermediate vector(s) which are output for an intermediate area 360 may be calculated using an area 362 which is larger than and encloses the area 360, or indeed an area which is smaller than the area 360. These methods shown in Figures 75 and 77  
30 minimise edge effects caused by small parts of larger objects extending into adjacent intermediate areas.

In the arrangement described with reference to Figures 1 to 48, the motion vector for a pixel is selected from the global vector(s) for the whole picture and the local vector(s) for the pixel under  
35 consideration. This scheme is expanded to include the possibility of selecting from one or more types of intermediate vector for intermediate areas including the pixel under consideration, as shown in



Figure 79. The local vectors 359 for the picture are used to form a global vector 362 for a 2 x 2 array of intermediate areas; second intermediate vectors 364 for a 3 x 3 array of intermediate areas; and third intermediate vectors 366 for a 9 x 6 array of intermediate areas. 5 The global, intermediate and original local vectors for each pixel are then supplied in combination to the motion vector selector 46 (Figure 4).

In the above description, reference has been made to acquiring progressive scan format signals from photographic film and to using 10 output progressive scan signals to record photographic film. It will be appreciated that other image sources may be used and other end products may be generated. For example, the input images may be computer generated, or produced by an animation camera, or video equipment.

Reference has also been made above to a 3232 pulldown format. It 15 will be appreciated that other pulldown formats, such as 3223, 2323, or 2332 may alternatively be used.

This application is one of a series of twelve applications filed on the same day and bearing agents' references N559-2 to N559-10 and N563-11 to N563-13. The content of each of the other applications is 20 incorporated herein by reference. (Application Nos. 90/24836.0, 90/24829.5, 90/24828.7, 90/24827.9, 90/24838.6, 90/24835.2, 90/24816.2, 90/24837.8, 90/24825.3 and 90/24817.0, 90/24826.1, 90/24818.8 respectively.)

CLAIMS:

1. A method of forming an output image by motion compensated interpolation between a pair of input images, comprising the steps, for  
5 each pixel location in the output image, of:  
    producing a respective motion vector;  
    determining respective source locations in the input images in dependence upon the respective motion vector; and  
    calculating the value of the respective pixel in the output image  
10 from respective values derived for the pixels at the source locations in the input images;  
    wherein the source locations are displaced from the location in the output image by amounts proportional to the respective motion vector plus a constant amount.  
15
2. A method as claimed in claim 1, wherein the constant amount is one quarter of a pixel location spacing in each dimension.
3. A method as claimed in claim 1 or 2, wherein the output image is  
20 interpolated half-way between the input images and the constants of proportionality are equal in magnitude and opposite in sign for the respective two source locations.
4. A method of forming an output image by motion compensated  
25 interpolation between a pair of input images, substantially as described with reference to Figures 61 to 64.
5. A method of forming a series of output images from a series of input images in which some of the output images are temporally aligned  
30 with respective input images and some of the output images are temporally offset from the input images, comprising the steps of:  
    producing a respective motion vector for each pixel location in each output image;  
    in the case of a temporally offset output image, determining, for  
35 each pixel location therein, respective source locations in the temporally preceding and temporally succeeding input images which are displaced from the pixel location in the output image by amounts

dependent upon the respective motion vector and the temporal offset;

in the case of a temporally aligned output image, determining,  
for each pixel location therein, a respective source location in the  
temporally aligned input image, and a respective source location in the  
5 preceding or succeeding input image which is displaced from the pixel  
location in the output image by an amount dependent upon the respective  
motion vector; and

calculating for each pixel location in each output image the  
value of the pixel from respective values derived for the pixels at the  
10 two source locations in the respective input images.

6. A method of forming a series of output images from a series of  
input images in which some of the output images are temporally aligned  
with respective input images and some of the output images are  
15 temporally offset from the input images, substantially as described  
with reference to Figure 65.

7. A method as claimed in any claims 1 to 3 and 5, wherein the  
values derived for the pixels at the source locations are derived by  
20 spatial interpolation between a set of pixel values around the source  
location.

8. An apparatus adapted to perform the method of any preceding  
claim.

**Patents Act 1977**  
**Examiner's report to the Comptroller under**  
**Section 17 (The Search Report)**

Application number 9024835.2

**Relevant Technical fields**

(i) UK Cl (Edition K ) H4F (FEP, FER, FGXX)

(ii) Int Cl (Edition 5 ) H04N, G06F

**Databases (see over)**

(i) UK Patent Office

(ii) Online Databases: WPI

**Search Examiner**

P J Easterfield

**Date of Search**

16 April 1991

Documents considered relevant following a search in respect of claims

1 to 8

Category (see over)	Identity of document and relevant passages	Relevant to claim(s)
	NONE	

SF2(p)

- 71 -

Category	Identity of document and relevant passages	Relevant to claim(s)

### Categories of documents

**X:** Document indicating lack of novelty or of inventive step.

**Y:** Document indicating lack of inventive step if combined with one or more other documents of the same category.

**A:** Document indicating technological background and/or state of the art.

**P:** Document published on or after the declared priority date but before the filing date of the present application.

**E:** Patent document published on or after, but with priority date earlier than, the filing date of the present application.

**&:** Member of the same patent family, corresponding document.

**Databases:** The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).

**THIS PAGE BLANK (USPTO)**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☒ **FADED TEXT OR DRAWING**

☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**